Phenotypic characterization and genetic variation of vitamin E genes in sunflower

by

Linchay Janine Daniels

submitted in accordance with the requirements for
the degree of

MASTER OF SCIENCE

In the subject

LIFE SCIENCES

at the

University of South Africa

Supervisor: Dr ZH Swanevelder

09 June 2017

Student number: **53024044**

I declare that: ***Phenotypic characterization and genetic variation of vitamin E genes in sunflower*** is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

SIGNATURE DATE 09 June 2017

Ms. Linchay Janine Daniels

ACKNOWLEDGEMENTS

First and foremost I would like to thank God for the strength, health and wisdom. I want to acknowledge my husband Anvor Daniels, son Zeke Daniels, parents Deweronie and Jacques Arnolds, sister Lu-allen Scheepers, brother Jefferson Keyser and friends for their constant support and encouragement during my period of study.

My utmost thanks to Dr. Shaun Reeksting for his assistance and training on the GC MS/MS and the availability of their laboratory for the phenotypic experiments on tocopherol profiling. I would also like to thank my fellow laboratoty collegues, Thuto Ntsowe and Elisha Pillay for their emotional and physical support during the period of my study.

I would like to extend my gratitude to Mr Tshilidzi Matamela, Mr Andrew Mokhele and their team from Agricultural Research Council-Grain Crop Institute (ARC-GCI) in Potchefstroom for their help during sample collection. Thank you to Mr Andrew Mokhele for his assistance with the linoleic acid analysis on the DA 7250 NIR analyzer. I also want to thank Ms Stephanie Cornelissen for her assistance with the bioinformatic analysis and Single nucleotide polymorphism (SNP) discovery.

Lastly I would like to thank the Agricultural Research Council (ARC) - Professional development program (PDP) and National Research Foundation (NRF) for their financial support.

Abstract

Linchay Daniels & Dirk (ZH) Swanevelder
Agricultural Research Council, Biotechnology Platform, Onderstepoort, 0110, Pretoria, South Africa
University of South Africa, Florida, 1709, Johannesburg, South Africa
Email: Swanevelderd@arc.agric.za

Sunflower (*Helianthus annuus*) consists of high levels of polyunsaturated fatty acids, making its oil susceptible to oxidation. Tocopherols can retard or prevented oxidation. The aim of this study is to determine the phenotypic tocopherol (vitamin E) composition and genetic diversity of the biosynthetic pathway genes. Seeds were characterized for fatty acid and tocopherol content. A positive correlation was found between oleic acid, *γ* (*r* =0.17) and *δ* (*r* =0.23) tocopherol but none between linoleic acid and all four tocopherol derivatives. Vitamin E gene homologues were identified and a concomitant pathway constructed, with genes of interest sequenced to determine their genetic variation. A sunflower gene database was developed for these genes and used to obtain 489 SNPs and 145 indels from the accessions evaluated. Only 139 of these SNPs were located in the exon regions of the gene candidates. These exon-based SNPs may influence tocopherol flow through possible enzyme structural modifications.

Keywords: sunflower, tocopherol, fatty acids, vitamin E homologues, *HPT*, *HPPD, TMT, TC*

Research outputs and awards

Award: Daniels LJ. NRF Scarce Skills Scholarship for MSc 2013-2014.

Daniels LJ and Swanevelder ZH (2014). Genetic variation of candidate genes in the vitamin E biosynthetic pathway in relation to oxidative stability and shelf life of sunflower oil. Agricultural Research Council, Professional Development Program (PDP) Conference (2014). *Poster presentation*. Received the ***Best MSc Poster Presentation*** award.

Daniels LJ and Swanevelder ZH (2014). Genetic variation of candidate genes in the vitamin E biosynthetic pathway in relation to oxidative stability and shelf life of sunflower oil. South African Society of Bioinformatics & South African Genetics Society (2014). *Poster presentation*.

Daniels LJ, Reeksting SB and Swanevelder ZH (2015). Determining the correlation between tocopherols and unsaturated fatty acids content in selected sunflower accessions. Agricultural Research Council, Professional Development Program (PDP) Conference (2015). *Oral presentation*.

Daniels LJ and Swanevelder ZH (2016). Genetic diversity within the vitamin E biosynthetic pathway of sunflower. Plant and animal genome conference xxiv (PAG), San Diego, California. *Abstract reviewed, accepted and published.*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

% – Percentage

°C – Degree Celsius

°C/min – degree Celsius per minute

ANOVA – Analysis of variance

ARC – Agricultural Research Council

ARC-GCI – Agricultural Research Council Grain Crops Institute

BAM – Binary Alignment/ Map

BHA – Butylated hydroxyanisole

BHT – Butylated hydroxytoluene

BLAST – Basic Local Alignment Search Tool

BLASTx – BLAST protein search

bp – base pair

BWA – Burrows Wheeler Aligner

cc/ml – cubic centimetre per millilitre

dATP – deoxyadenosine triphosphate

dCTP – deoxycytidine triphosphate

dGTP – deoxyguanosine triphosphate

DNA – Deoxyribonucleic acid

dTTP – deoxythymidine triphosphate

E. coli – *Escherichia coli*

EST – Expressed sequence tag

g – grams

g/L – grams per litre

GATK – Genome Analysis Toolkit

GB – Gigabase

GC – Guanine cytosine

GC MS/MS – Gas chromatography Mass spectrometry & Gas chromatography Tandem Mass Spectrometry

GGDP – Geranylgeranyldiphosphate

GGP – Geneseek Genomic Profiler

GGR – Geranyl geranyl reductase

HGA – Homogentisate

HPP – p-Hydroxypyruvic acid

HPPD – p-Hydroxyphenylpyruvate dioxygenase

HPT – Homogentisate prenyltransferase

HPPT - Homogentisate phytylprenyltransferase

Id – identifier

IDT – Integrated DNA Technologies

IGV – Integrative Genomic Viewer

Kcal – kilocalorie

KEGG – Kyoto Encyclopedia of Genes and Genomes

L – Lipid radical

LOO – Peroxy radical

LOOH – Lipid hydroperoxide

m – meter

MAS – Marker assisted selection

mg – milligram

$MgCl_2$ – Magnesium chloride

min – minute

ml – millilitre

ml/min – millilitre per minute

mm – millimetre

MRM – Multiple reaction monitoring

mRNA – Messenger RNA (ribonucleic acid)

MSBQ/PrBQMT – 2-Methyl-6-phytylbenzoquinol methyltransferase

NCBI – National Center for Biotechnology

NGS – Next Generation Sequencing

NMR – Nuclear magnetic resonance

NS – No significance

OH – Hydroxide

PCR – Polymerase Chain Reaction

PDP – Phytyl diphosphate

PDS1 – phytoene desaturation

PE – Pair end

PK – Phytol kinase

ppm – parts per million

PQ – Plastoquinone

QTL – Quantitative trait loci

R1 – Reproductive stage (terminal bud forms)

R5.1 – Beginning of flowering (10%) are subdivided depending on the completion of the head area or is flowering e.g. 5.3 (30%)

RAPD – Random amplified polymorphic DNA

RFLP – Random fragment length polymorphism

rmp – revolutions per minute

RNA – Ribonucleic acid

s – second

SAM – Sequence Alignment/Map

SLK – Spinlock nuclear resonance

SNP – Single nucleotide polymorphism

SR – Standard reference

SRS-EMBL – Sequence Retrieval System – European Molecular Biology Laboratory

SSR – Simple sequence repeat

TAE – Tris(hydroxymethyl)aminomethane, acetic acid and Ethylenediaminetetraacetic acid

TAIR – The *Arabidopsis* Information Resource

TBHQ – tert-butyl hydroquinone

tBLASTn – protein BLAST to translated nucleotide database

TC – Tocopherol/tocotrienol cyclase

TE – Tris and EDTA (Ethylene Diamine Triacetic Acid)

TIGR – The Institute of Genomic Research

TM – melting temperature

TMT – Tocopherol/tocotrienol methyltransferase

Uniprot – Universal protein resource

USDA – United States Department of Agriculture

V – voltage

V2 – Second leaf (vegetative stage)

VCF – Virtual Contact File

VE – Vegetative Emergence (seedling has emerged)

α – Alpha

β – Beta

γ – Gamma

γ-T3 – gamma-Tocopherol 3

δ – Delta

µl – microliter

µM – micro molar

Chapter 1

Introduction

1.1 Background and rationale

Sunflower (*Helianthus annuus* L.) belongs to the family Asteraceae and the genus *Helianthus* that comprises of 68 known species of which only two are used for food purposes (e.g. *H. annuus* for oil, birdfeed or animal feed and *H. tuberosus* used as a root vegatable) (Heiser 1978, Shahidi 2005). The genus name *Helianthus* is derived from the Greek word *helios* which means sun and *anthos* meaning flower (Shahidi 2005). Sunflower is an important oil seed crop that produced oil that mainly consists of linoleic and oleic acid (about 90%) (Škorić et al 2008, Awatif and Shaker 2014). Vegetable oils with high levels of polyunsaturated fatty acids are highly susceptible to oxidation, making oil oxidative unstable and with a short life span (Mermelstein 2010). Oxidation can be retarded or prevented by the presence of antioxidants, such as tocochromanols. Tocochromanols (tocopherols and tocotrienols), also known as vitamin E, is a collective name for a group of fat-soluble compounds with distinct antioxidant properties protecting cells from oxidative damage (Herrera and Barbas 2001). These fat-soluble compounds comprise of various tocopherol and tocotrienol derivatives (Hofius and Sonnewald 2003, Rengaraj and Hong 2015). Since sunflower oil consists of high levels of polyunsaturated fatty acid, it is highly susceptible to oxidation. This becomes a problem when sunflower seed are harvested, oil extracted and stored for a long period of time. The presence of high levels of vitamin E in oil would hamper the process of oxidation resulting in a longer shelf life.

It is therefore important to understand the vitamin E content of sunflower lines and the genetic variation in the specific genes involved in the biosynthetic pathway of vitamin E production to be able to associate genetic polymorphisms in relation to variation found in vitamin E content. This study focuses on determining both the phenotypic composition of tocopherols within

sunflower lines as well as the genetic diversity of four genes involved in the biosynthetic production pathway of vitamin E.

Presently, the vitamin E biosynthetic pathway is unknown for sunflower, but known in other plants. Additionally, the genetic diversity of these genes is also unknown in sunflower. The lack of this crucial information is preventing the incorporation of superior vitamin E traits in the ARC sunflower-breeding program. To successfully incorporate beneficial vitamin E traits into the sunflower-breeding program we need to identify the tocopherol levels and associated mutations amongst the accessions within this resource and thus identify possible lines for further breeding processes. This will be determined using the high-throughput next generation sequencing gene target approach (Marroni et al 2011). This technology can be used to detect large populations of rare functional variants in pooled samples (Druley et al 2009, Out et al 2009). Fritsche et al 2012 used the candidate gene-based approach to identify and assess the role of polymorphisms in *B. napus* tocopherol biosynthesis genes on tocopherol content and composition and found this approach a very effective method for SNP identification. This argues for the suitability of this approach in determining genetic polymorphisms in sunflower genotypes on the biosynthetic pathway of vitamin E. Four genes will be targeted using PCR primers designed from the vitamin E homologues identified via bioinformatics, the amplicons will be randomly fragmented before next generation sequencing (NGS) (Marroni et al 2011). The NGS technology generates large amounts of genetic data at a cost effective level (Marroni et al 2011).

1.2 Problem identification

The successful incorporation of the vitamin E traits in sunflower breeding programs will rely on understanding the level of genetic polymorphisms among the specific genes in the vitamin E pathway and its distribution across the germplasm accessions within the collection. Presently, the genetic variation associated with vitamin E in the sunflower accessions, and thereby the economic traits of oxidative stability and shelf life, are unknown within the

germplasm collection of the ARC. Understanding the diversity of these genes and their association with vitamin E (and its linked traits), will help the breeding program in identifying possible parental lines with the desired traits, as well as provide possible polymorphisms for marker assisted tracking of these traits during hybrid development. This crucial information will aid in the incorporation of superior economic traits (e.g. oil stability and shelf life) within sunflower breeding program. In this work, the vitamin E biosynthetic pathway is investigated and elucidated from literature using gene homologues and the draft sunflower genome. Selected lines within the ARC sunflower germplasm collection were also phenotyped for their oil tocopherol content and based on these results, a subsection were screened with the identified vitamin E biosynthetic gene candidates, to investigate their diversity. For the purposes of this work, assessment of genetic polymorphisms will be limited to the four major genes involved in the biosynthetic pathway of vitamin E.

1.3 Aims

The aims of the project are to determine the phenotypic tocopherol composition and genetic diversity of vitamin E genes in the biosynthetic pathway of tocochromanols. This will be done to link vitamin E genotype and production and thereby, through association, to economically important traits like oxidative stability and shelf life.

1.4 Objectives

1. To determine the tocopherol phenotypes of selected sunflower accessions in the ARC's germplasm collection.

2. To identify the vitamin E biosynthetic gene homologues in the sunflower draft genome and public bioinformatic available resources and to construct a vitamin E biosynthetic pathway for sunflower.

3. To amplify the identified gene homologues with custom designed primers in a subset samples selected based on their phenotypic profiles identified from the sunflower germplasm collection.

4. To determine the polymorphisms of these genes though a next generation sequencing (NGS) approach.

5. To determine the genetic diversity of the candidate genes through single nucleotide polymorphism (SNP) calling.

References

Awatif I and Shaker M 2014. Quality characteristics of high-oleic sunflower oil extracted from some hybrids cultivated under Egyptian conditions. *Helia* 113-126

Druley TE*,* Francesco LM, Vallania FL, Wegner DJ, Varley KE, Knowles OL, Bonds JA, Robison SW, Doniger SW, Hamvas A, Cole FS, Fay JC, and Mitra RD 2009. Quantification of rare allelic variants from pooled genomic DNA. *Nature Methods* 1-6

Fritsche S, Wang X, Li J, Stich B, Kopisch-Obuch FJ, Endrigkeit J, Leckband G, Dreyer F, Friedt W, Meng J and Jung C 2012. A candidate gene-based association study of tocopherol content and composition in rapeseed (*Brassica napus*). *Frontiers in Plant Science* 1-16

Heiser CB Jr. 1978. Taxonomy of *Helianthus* and Origin of Domesticated Sunflower. In: Sunflower Science and Technology. *American Society of Agronomy, Crop Science Society of America, Soil Science Society of America* 31-53

Herrera E and Barbas C 2001. Vitamin E: action, metabolism and perspectives. *Journal of Physiology and Biochemistry* 43-56

Marroni F, Pinosio S, Di Centa E, Jurman I, Boerjan W, Felice N, Cattonaro F and Morgante M 2011. Large-scale detection of rare variants via pooled multiplexed next-generation sequencing: towards next-generation Ecotilling. *The Plant Journal* 736-745

Mermelstein NH 2010. Improving soybean oil. *Food safety and quality* 72-76

Škorić D, Jocic S, Sakac Z and Lecic N 2008. Genetic possibilities for altering sunflower oil quality to obtain novel oils. *Canadian Journal of Physiology and Pharmacology* 215-220

Shahidi F 2005. Bailey's industrial oil and fat products. In: Sunflower Oil. Shahidi F and Grompone MA 665-725

Out AA, van Minderhout IJHM, Goeman JJ 2009. Deep sequencing to reveal new variants in pooled DNA samples. *Human Mutation* 1703-1712

Chapter 2

Literature review

2.1 Introduction

Sunflower (*Helianthus annuus* L.) belongs to one of the largest plant families the Asteraceae, which consists of approximately 24 000 described species (Heiser 1978, Stevens 2001). Native to the Americas (Hewezi et al 2006), sunflower has been exploited since the early 16th century as food in mixtures of cooked vegetables and in food "concentrates" by some native Americans (Putt 1978, Janick 2013). It was introduced to Spain in the middle of the 16th century and has since spread worldwide where it is grown extensively as one of the most important oil seed crops of the world (Putt 1978, Berglund 2007).

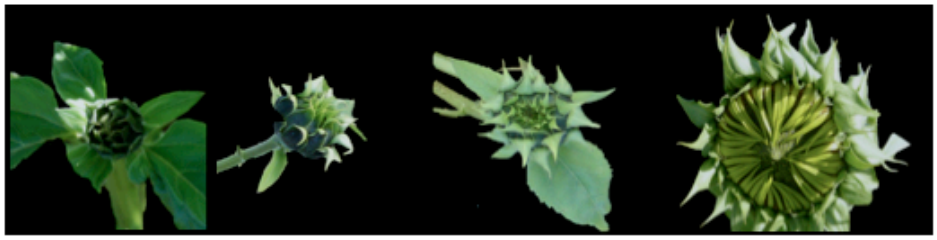
Sunflower is regarded as one of the most promising oil seed crops when it comes to the genetic alteration of oil quality (Scharp 1986, Škorić et al 2007). The first significant alteration of sunflower oil composition was done by Soldatov in 1976 (Škorić et al 2007). He used chemical mutagenesis to obtain a genotype that produced high oleic acid containing seeds. This mutant was developed and released as the cultivar 'Pervenet', a line that produced seeds that contained 85-90% oleic acid (Škorić et al 2007). Later in the 1980s more efforts were made to select sunflower genotypes with higher oil content (Fernández-Martínez et al 1989). Some of the first improvements took place in Russia on trait selection and hybridization aiming to increase oil content of the seed (Iuoras et al 2002). A further boost to sunflower breeding was made possible by the discovery of the male–sterile and restorer gene system (Vranceanu and Stoenescu 1971, Schnable and Wise 1998). Utilizing this system led to the production of new inbred lines and further increased commercial interest in the crop (Leeuwner 2005).

Sunflower is one of only a few oil crops produced in South Africa, and as a grain crop ranks third in production after maize and wheat (Department of

Agriculture Forestry and Fishery 2010). Sunflower is capable of growing under hot and dry conditions, thus enabling planting in eight of the nine provinces, with 85% of the total harvest coming from the Free State and the North West provinces (Department of Agriculture Forestry and Fishery 2010). In South Africa, sunflower is grown as a summer rainfall crop and can be planted from the beginning of November to early January. Various studies found that if planting is delayed (by between 20 - 30 days), the head diameter, seed yield and oil content decrease (Miller et al 1984, Dedio 1985, Killi and Altenbay 2005). When planted during planting season the oilseed concentration will only start to increase a few days after flowering and stops a few days before physiological maturity (Goffner et al 1988).

Sunflower is a fast growing crop that undergoes a number of vegetative and reproductive growth stages before it reaches maturity (Berglund 2007). In short the first growth stages are the emergence and expansion stages (10-20 days), also referred to as the VE stages (±10 days), which show the appearance of the hypocotyl, ending just before the first pair of leaves reaches 4cm in length (±20 days). After about 20 days the second set of growth stages occurs, collectively known as the vegetative stage (V2, V4, V8, and V20) (figure 2.1) and takes place between 20 to 40 days. The third set of stages, referred to as the reproductive stage, includes various growth developmental stages, like the emergence of the flowering bud (R1 to R4) (figure 2.1), which occurs between 46 to 71 days. This is followed by the flowering (R5.1, R5.2, and R6) and seed production stages (R7, R8 and R9) (figure 2.1) which occur between 72 and 120 days, where seed harvesting can occur after 120 days (Berglund 2007). Sunflower reaches maturity between 90 and 120 days after sowing, depending on the genotype and growth conditions (Berglund 2007).

V-E V-2 V-4



R1 R2 R3 Top view R4 Top view

R-5.1 R-5.5 R-5.9 R-6

R-7 R-8 R-9

Figure 2.1: The developmental stages of sunflower: VE represents the emergence and expansion stages, which are followed by the vegetative (V2,

V4, V8, and V20), flower emergence (R1 to R4), flowering (R5.1, R5.5, R5.9 and R6) and seed production stages (R7, R8 and R9).

Sunflower is the most important oilseed crop and the largest source of vegetable oil in South Africa, followed by soybean and canola oil (van der Merwe et al 2013). Sunflower oil production has contributed tremendously towards growth of the South African agriculture over the years. In the period between 2000 and 2009 an average of 682 thousand tons of sunflower seeds were produced with a gross value of R1.4 billion per annum (Department of Agriculture Forestry and Fishery 2010). Between 2005 and 2009 North West province produced the most sunflower seed, but production has been declining since. In ten years ending in 2010, South Africa produced about 230 000 tons of sunflower oil, representing about 30% of the seed produced within this time frame (Department of Agriculture Forestry and Fishery 2010). Sunflower producers received about R1000 per ton before the 2000/2001 growing season, but the price has consistently increased, with R4857 per ton achieved in 2013 (Department of Agriculture Forestry and Fishery 2013) and R5015 per ton in 2015 (Department of Agriculture Forestry and Fishery 2015). The market price in 2016 fluctuated due to drought, but prices in excess of R6800 per ton were achieved (National Agricultural Marketing Council 2016). This indicates that sunflower contributes tremendously to the South African economy.

The quality of sunflower oil refers to the seed oil content and degree of saturated and unsaturated fatty acids within it (Kilcast and Subramaniam 2000). Sunflower oil consists of high levels of polyunsaturated fatty acids making it easily susceptible to oxidation (Aluyor and Ori-Jesu 2008). The presence of oxygen and high temperatures causes sunflower oil to turn rancid, resulting in unstable oil with short life spans (Tavassalkar et al 2012). The oxidation of sunflower oil can be prevented by the presence of vitamin E, e.g. tocopherols and tocotrienols, also referred to as antioxidants (Tavassalkar et al 2012). The unavailability of antioxidants, like vitamin E, will reduce the commercial quality of the product. Therefore it is important to understand the degree of fatty acid saturation and vitamin E content, and how

it is affected by temperature and oxidation, to be able to breed for economic traits, like oil stability and shelf life.

2.2 Sunflower fatty acids and oil quality

South Africa can be regarded as a liquid oil market since cooking oil accounts for 25-30% of the total oils and fats consumption in the country, with sunflower oil as the dominant oil source (MPOC 2009). Indeed, sunflower oil contribution is estimated to represent around 80% of vegetable oil utilized within South African's diets (van der Merwe et al 2013). Sunflower is highly suitable in the food industry because of its lack of odour and colour but the most prominent characteristic is its high levels of polyunsaturated fatty acids (Harris et al 1978, Sadoudi et al 2014). It is used as suitable cooking oil because of its neutral taste (Soyatech foods, growing opportunities, http://www.soyatech.com/). The quality of sunflower oil, as already mentioned, is related to seed oil content and fatty acid composition, which defines the value of the oil for industry use (Rondanini et al 2003).

Sunflower oils consist of several types of fatty acids, which include saturated acids, e.g. stearic and palmitic acids, monounsaturated acid, e.g. oleic acid, and polyunsaturated acid, e.g. linoleic acid (Lacomble and Berville 2001, Awatif and Shaker 2014). Acceptable linoleic and oleic acids' percentages ranges between 48 – 75 and >80%, respectively, for sunflower oil (Lacomble and Berville 2001). High levels of fatty acid and the levels of tocopherols, sterols, carotenoids, and other compounds, all contribute to sunflower oil quality (Awatif and Shaker 2014). The characterization of sunflower oil is usually done by its oleic acid composition, i.e. mid-oleic (55-75%), oleic (>80%) or high oleic acid (80-95%), depending on the predominant type of fatty acid (Aluyor and Ori-Jesu 2008). Typical sunflower oil with linoleic amino acid content usually contains about 65% polyunsaturated fat, 21% monounsaturated fat and 11% saturated fat (National sunflower association http://www.sunflowernsa.com/oil/linoleic-sunflower-oil/, May 2014).

The two polyunsaturated fatty acids, linoleic (C 18:2) and oleic acid (C 18:1), are 18 carbon fatty acid with two or one double bonds, respectively, and together make up about ±90% of the total higher fatty acid content in conventional sunflower oil (Škorić et al 2008, Awatif and Shaker 2014). The relation between oleic and linoleic acid is strongly affected by environmental temperature during seed development, therefore it is difficult to produce oil with a constant degree of unsaturation (Škorić et al 2008). Kovacik et al 1998 found that seeds which mature under high temperatures produce oil high in oleic acid and low in linoleic acid compared to those produced during low temperatures. The saturated fatty acids, stearic (C 18:0) and palmitic acids (C 16:0) together account for ±8-10% of the total fatty acid content (Gunstone 2002, Awatif and Shaker 2014). In addition, there are trace amounts of other higher fatty acids namely, myristic (C 14:0), myristoleic (C 14:1), palmitoleic (C 16:1), arachidic (C 20:0) and behenic acids (C 22:0) (Friedt et al 1994, Awatif and Shaker 2014). These unsaturated fatty acids (oleic and linoleic acid) are considered essential because the human body cannot synthesize them (Abitogun and Oshodi 2010).

2.3 Oxidative stability and shelf life

Lipid oxidation is one of the basic chemical reactions that causes food deterioration, thereby resulting in undesirable food products (Babovic et al 2010). The term oxidation is defined as the interaction between oxygen molecules and all the different substances they interact with (McLaren 2014). During the oxidation process, at least one electron is lost when two or more substances interact, thus producing a free radical. Free-radicals are chemicals that contain one or more unpaired electrons, which make them highly reactive due to the presence of this unpaired electron (Colombo 2010, Kabel 2014). The free-radical chain reaction in the oxidation process involves three steps namely, initiation, propagation, and termination (figure 2.2). According to O'Connor and O'Brien (2006) the initiation step is the key factor in maintaining the quality of food products and extending its shelf life. Indeed, in the autoxidation of unsaturated fatty acids it is the initiation step where the formation of free-radicals is formed. This step may be influenced by factors

such as irritation, enzymes or active oxygen species (O'Connor and O'Brien 2006). The reaction is usually activated by the removal of hydrogen from the methylene group, where the resulting free radical then reacts with oxygen to form peroxide free-radicals (O'Connor and O'Brien 2006). This peroxide free radical reacts with another unsaturated molecule, where the chain of reaction continues to create a hydroperoxide (O'Connor and O'Brien 2006).

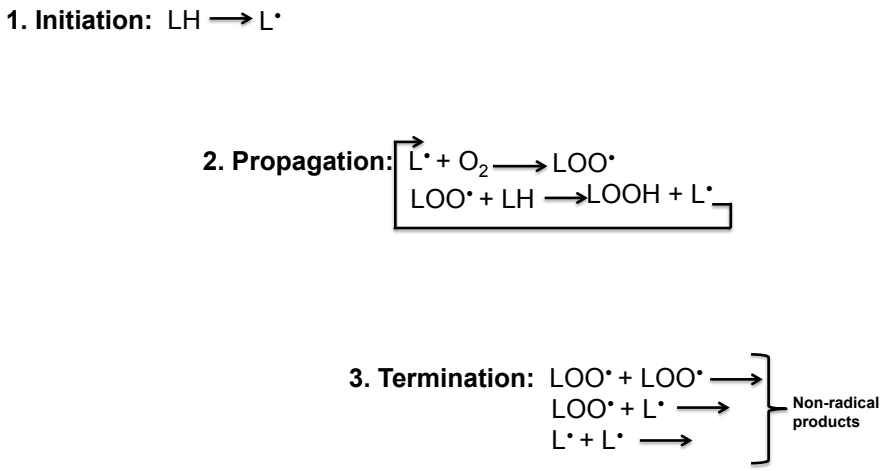


Figure 2.2: **The process of lipid oxidation:** Lipid oxidation process consists of initiation, propagation and termination in which a lipid radical (L) forms during the initiation step. This process continues to the propagation step where the lipid radical reacts with atmospheric oxygen, resulting in peroxy-radical (LOO). The propagation step may also occur by removing hydrogen from another acyl chain resulting in a lipid hydroperoxide (LOOH) and a new radical (L). The termination step occurs when two free-radicals react to form a non-radical product (O'Connor and O'Brien 2006).

Vegetable oils (e.g. soybean and sunflower) that contain high levels of polyunsaturated fatty acids are health beneficial but are highly susceptible to oxidation when exposed to very high temperatures, resulting in horrible odour, flavour and shorter shelf life (Mermelstein 2010). The shelf life of oil is based

on its chemical stability making it directly dependent on the oxidative stability of the oil (Aluyor and Ori-Jesu 2008). High temperature uses of sunflower oil, such as frying, need oils that are highly resistant to thermo-oxidation, i.e. usually oils that have high concentrations of antioxidants (*γ* and *δ* tocopherols) and oleic acid (Fernández-Martínez et al 2009). Indeed, studies have shown that there is a positive relation between tocopherol content and polyunsaturated fatty acids (Herting and Drury 1963, Kamal-Eldin and Anderson 1997, Goffman and Bohme 2001). Karlovic et al (1997) demonstrated the importance of the tocopherol and oleic acid in oil stability by performing the Rancimat test at 100°C on oil from different sunflower genotypes. The Rancimat test is used for automatic determination of oxidative stability of natural oils, fats or fat containing products (Méndez et al 1996, Ciemniewska-Żytkiewicz et al 2014). Results obtained from the study of Karlovic et al (1997) revealed that common sunflower oil could remain stable for only 8 hours, while high oleic acid sunflower oil, with 50 % *β* tocopherol, could be oxidatively stable for up to 168 hours. Oleic-type sunflower oil was only stable for 33 hours, whereas the oleic-type with high *γ* tocopherol content remained stable for up to 152 hours. Indeed, the oxidative stability of vegetable oil depends on its fatty acid composition, thus oil with high oleic acid content has greater shelf life and oxidative stability than sunflower oil with high linoleic acid content (Warner et al 1997).

2.4 Tocochromanols as antioxidants

Vitamin E was discovered in 1922 by Evans and Bishop, but Epstein et al only described its antioxidant function in 1966 (Epstein et al 1966). As already mentioned, the antioxidant activity of vitamin E is due to its ability to donate phenolic hydrogen to lipid free-radicals on the chromanol head (which contain a phenolic and heterocyclic ring) (Kamal-Eldin and Appelqvist 1996, Lobo et al 2010). These antioxidants form a hydroperoxide and tocopheroxyl radical by donating a hydrogen atom to peroxyl radicals of unsaturated lipid molecules (Romero et al 2004, Lobo et al 2010). Free-radicals with their unpaired electrons can react with most molecules and result in disruption of cell functionality. All derivatives of tocochromanols are able to reduce free-

radical damages by breaking down lipid peroxidation, protecting the cell membrane by lipid repair and replacement (Kabel 2014). Tocochromanols also protects the cells by neutralizing free-radicals before they can react and cause DNA damage or lipid oxidation (Colombo 2010). The relative antioxidant activity of *α, β, γ* and *δ* tocochromanols depend on composition of the tocopherols, temperature, physical state and lipid composition (Lea and Ward 1959, Khan et al 2015). According to Azizkhani et al (2011) tocopherols are the most important lypophilic natural antioxidants. The oxidative stability of oil is not only intrinsically related to the degree of unsaturated oils (Jorge et al 1996), but its stability may also be improved by tocopherols that share an interaction with phospholipid (Ikeda and Fukuzumi 1977, Hamilton et al 1998, Bandarra et al 1999, Shahidi et al 2013). Tocopherols and phospholipids show a synergistic effect for enhancing the oxidative stability of oils, therefore optimizing the concentrations of tocopherols may be affected by the amount of phospholipids and other minor compounds (Jung and Min 1990, Hamilton et al 1998, Shahidi et al 2013). Phospholipids are made up of oxygen and phpsohorus (that contains a phosphate group and two fatty acids attached to one glycerol) and can alter the structure and function of lipids. It can also act as an antioxidant through their metal scavenging, catalytic and synergistic activity (Carelli et al 1997).

All tocochromanol derivatives have an antioxidant activity but tocopherol is believed to play additional roles with its antioxidant feature (Falk and Munné-Bosch 2010). According to Sattler et al (2004) tocopherols play a major role in seed longevity by reducing the accumulation of lipid oxidation during storage. Tocopherols also play a role in controlling redox homeostasis and gene expression by modifying and controlling the levels of lipid peroxidation in the leaves (Sattler et al 2006, Falk and Munné-Bosch 2010). It is also involved in intracellular signalling by regulating the amount of jasmonic acid in the leaves (Boscoboinik et al 1991, Rimbach et al 2002, Munné-Bosch et al 2007). As an antioxidant, tocopherols and tocotrienols attribute to the inhabitation of membrane lipid peroxidation, controlling cell damage that is caused by free-radicals (Liebler 1993, Hunter and Cahoon 2007, Colombo 2010).

The critical minimal concentrations of antioxidants are usually associated with the inhibition of vegetable oil oxidation (Márquez-ruiz et al 2003). When tocopherols are added to triacylglycerols, the activation energy of antioxidants increases and its degradation is restrained. Triacylglycerols are esters joined together by one glycerol and three fatty acids and are stored as fat energy in adipose tissues. Triacylglycerols that contain unsaturated fatty acids consist of more liquid due the presence of kinks in unsaturated oils. In studies conducted on the antioxidant properties of *α* and *γ* tocopherols of rapeseed oil triacylglycerols at 40°C, it was found that *α* tocopherol was a more stable antioxidant at concentrations ≤50 ppm than *γ* tocopherols, but at concentrations of 100 ppm, *γ* tocopherols appeared more efficient and stable as an antioxidant (Lampi et al 1999).

Fuster et al (1998) worked on purified sunflower triacylglycerols which were oxidized at 55°C and found that at 50 ppm *α* tocopherol was more efficient antioxidant than *γ* tocopherol, but *α* tocopherol looses efficiency at 200 ppm. Corsini et al (2009) studied the tocopherol and oxidation loss during frying temperatures at 180°C for 25 hours and found that the smallest changes occurred in palm oil (which is mostly a saturated oil), but indicated a clear decrease in both tocopherol concentration and oxidation stability in sunflower and cottonseed oil (which is more unsaturated oil). Corsini et al (2009) concluded that the minimum loss in stability and tocopherol concentrations was found in the oil with the lowest concentrations of unsaturated fats. These studies underline the importance of tocopherols and its antioxidant properties. Other works concluded that antioxidants are needed by a plant to be able to prevent or retard oxidation deterioration and to increase and extend the lifespan of many products (Abdalla and Roozen 1999, Zandi and Gordon 1999, Codex Alimentarius Commissions 2001, O'Brien 2004, Robbins and Sewolt 2005).

2.5 Vitamin E (Tocochromanols)

Vitamin E, discovered by Evans and Bishop in 1922, was initially a collective name used for a group of fat-soluble compounds with distinct antioxidant

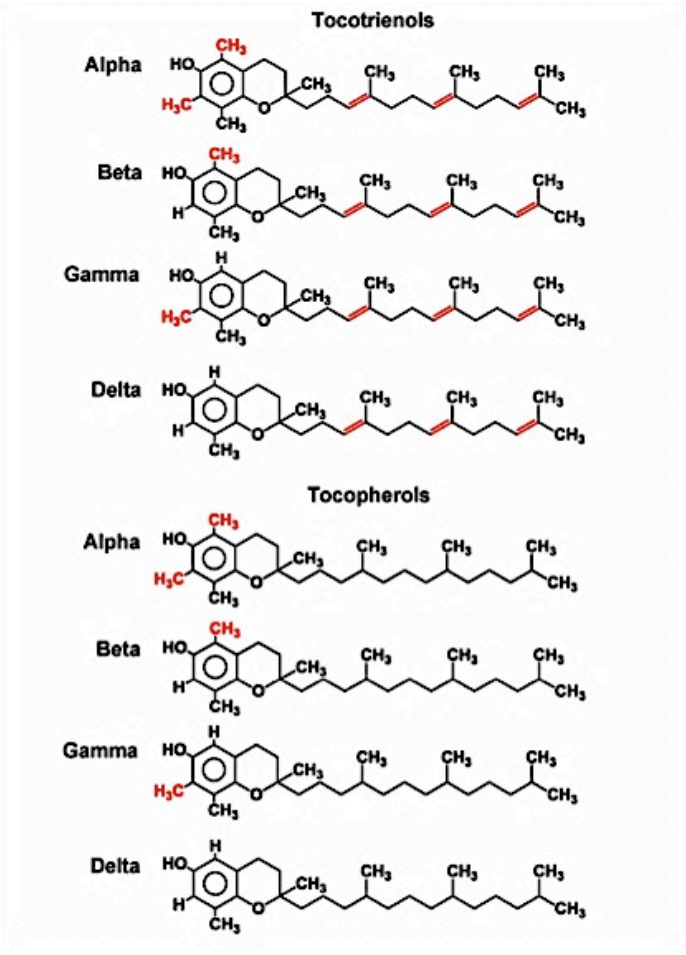
properties (Herrera and Barbas 2001, Rizvi et al 2014). These fat-soluble factors were first tested on rats and were found to be essential for their reproduction (Mackenzie et al 1939, Rengaraj and Hong 2015). They are also important in humans – being the prime defender in the body against oxidation, i.e. preventing oxidation of polyunsaturated fatty acids (Khan et al 2015). It was shown that Vitamin E intake reduces heart diseases, cancer, cardiovascular disease, diabetes and infections in humans (Aggarwal et al 2010). Among the fat-soluble factors comprising Vitamin E are various tocopherol and tocotrienol derivatives (Hofius and Sonnewald 2003, Rengaraj and Hong 2015).

The combined name for tocopherols and tocotrienols is tocochromanols (Hunter and Cahoon 2007, Rizvi et al 2014). The word tocopherol is derived from the Greek word "*tokos*" which mean "childbirth" followed by "*phorein*" meaning "bear" and "*ol*" which was added for phenolic nature (Kamal-Eldin and Appelqvist 1996, Zielinska and Nowak 2014). Tocotrienols are named by analogy to tocopherols also meaning, "to bear" or "pregnancy" but they are also known as "trienes" due to structural differences. The tocotrienol hydrocarbon tail contain three trans double bonds at the 3', 7' and 11' positions, while tocopherols consist of a fully saturated hydrocarbon tail (figure 2.3) (Hunter and Cahoon 2007, Zielinska and Nowak 2014).

There are four natural types of tocopherol and tocotrienols, referred to as *α, β, γ,* and *δ*, which only differ by the number and position of methyl groups on the aromatic rings (figure 2.3) (Hass et al 2006, Zielinska and Nowak 2014). *α*-Tocochromanols consist of three methyl groups on their aromatic ring, with *γ* and *β* tocopherols/tocotrienols having two methyl groups and *δ* tocopherol/tocotrienol containing only a single methyl group on their aromatic ring (Hunter and Cahoon 2007, Zielinska and Nowak 2014) (figure 2.3). It is the presence of the methyl groups on the chromanol head that is important in vitamin activity: an increase in methylation results in an increase of vitamin activity (Burton and Ingold 1981, Rimbach et al 2002). *α*-Tocopherol is considered to have more vitamin activity than the rest of the tocochromanols due to the higher number of methyl groups on its chromanol head (Baldwin et

al 1974, Raiola et al 2015). The four main components, i.e. *α* (5,7,8-trimethyl), *β* (5,8-dimethyl), *γ* (7,8-dimethyl) and *δ* (8-methyl), are synthesized by plants and other oxygenated organisms (Hunter and Cohoon 2007, Zielinska and Nowak 2014).

Tocopherols are present in all photosynthetic organisms while tocotrienols are only found in certain plant groups – exclusively in their seeds and fruits (Falk and Munné-Bosch 2010). *α*-Tocopherol is mainly found in the leaves of plants, i.e. specifically in the chloroplast of the plant cell, while *β, γ*, and *δ* homologues are usually found outside the chloroplast (Alscher and Hess 1993, Falk and Munné-Bosch 2010). According to various studies, *α* tocopherol is not only the primary form of tocopherols in leaves, but is also found in photosynthetic prokaryotes such as *Synechocystis* when compared to the rest of the tocochromanols, and it can be readily absorbed and retained by cells (Kamal-Eldin and Appleqvist 1996, Hofius and Sonnewald 2003, Hunter and Cahoon 2007).



Aggarwal et al (2010)

Figure 2.3: **The chemical structure of tocopherol and tocotrienols:** The only differences are three double carbon bonds in tocotrienols at position 3', 7' and 11', that do not occur in tocopherols (Aggarwal et al 2010, Falk and Munné-Bosch 2010).

2.6 Biosynthetic pathway of Tocochromanols

Molecular and biochemical studies show that tocopherol and tocotrienol share the same biosynthetic pathway (Schultz et al 1991, Falk and Munne´-Bosch 2010). The tocochromanol pathway is derived from the shikimate pathway and the hydrophobic tail is from the non-mevalonate pathway (Hofius and Sonnewald 2003). The first step of vitamin E biosynthesis occurs via the transfer of a prenyl group to form homogentisate. The tocochromanols head group proceeds via the condensation of homogentisate, with tyrosine as the basic precursor. Tyrosine is oxidized to p-hydroxypyruvic acid (*HPP*) that is converted to homogentisic acid via the enzyme p-hydroxyphenylpyruvate

dioxygenase (*HPPD*). Homogentisic acid is concentrated with phytyl diphosate in a reaction by prenyl transferase. The action of homogentisate prenyltrasferase (*HPT*) increases the yields of 2-methyl-6-phytylbenzoquinol and 2-methyl-6-geranylgeranylbenzoquinol that is known as the common precursors of all tocopherols and tocotrienols (Hunter and Cahoon 2007). These compounds are further methylated by the enzyme 2-metyl-6-phytylbenzoquinol methyltransferase (*PrBQMT*) to form 2,3-dimethyl-5-phytylbezoquinol and 2,3-dimethyl-5-geranylgeranylbenzoquinol. This methylation step is excluded in the production of *β* tocochromanols but includes the second methylation step (Hunter and Cahoon 2007).

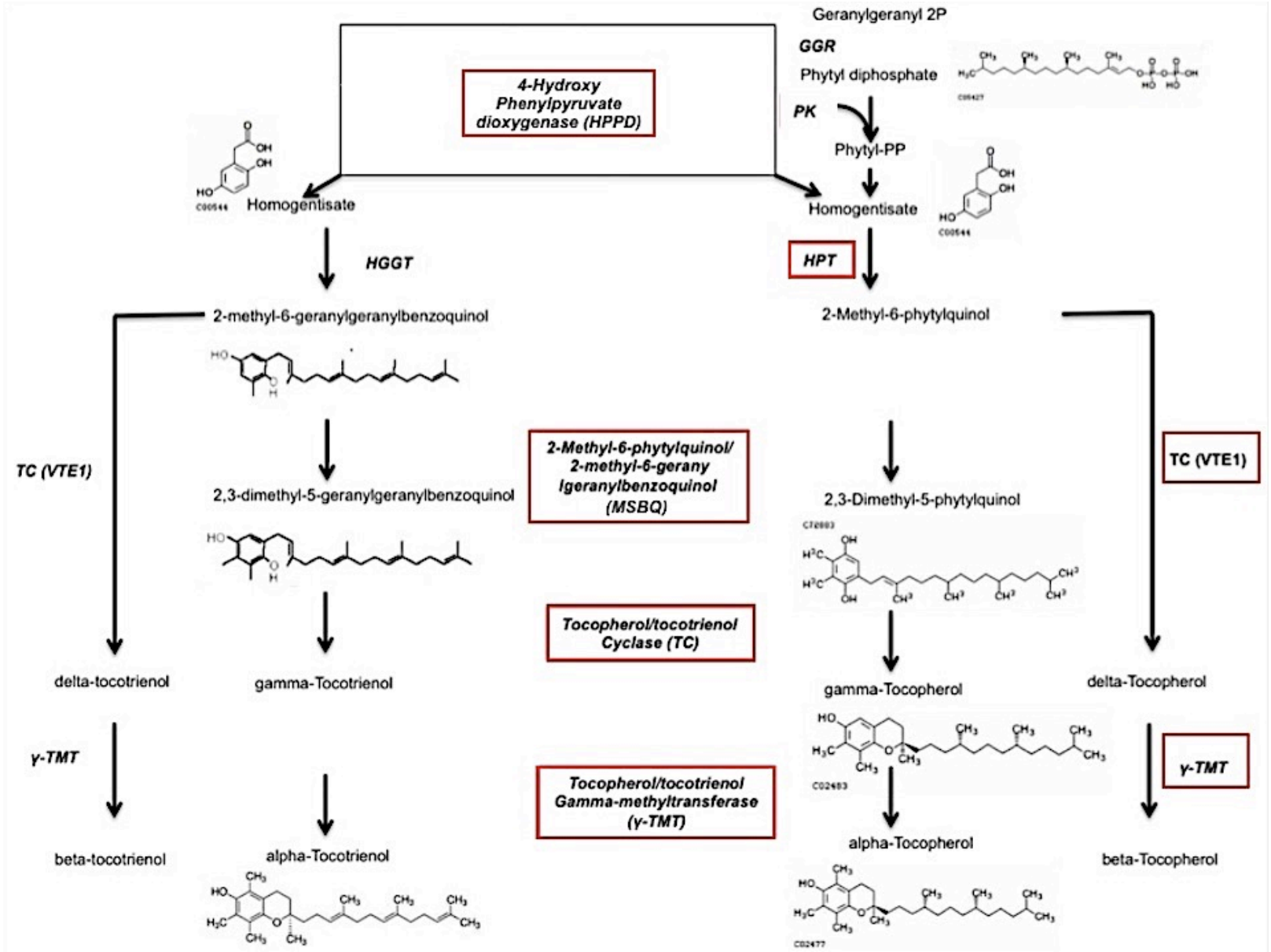

Geranylgeranyl 2P
GGR
Phytyl diphosphate
PK
Phytyl-PP
4-Hydroxy Phenylpyruvate dioxygenase (HPPD)
Homogentisate
Homogentisate
HGGT
HPT
2-methyl-6-geranylgeranylbenzoquinol
2-Methyl-6-phytylquinol
TC (VTE1)
2-Methyl-6-phytylquinol/ 2-methyl-6-gerany lgeranylbenzoquinol (MSBQ)
TC (VTE1)
2,3-dimethyl-5-geranylgeranylbenzoquinol
2,3-Dimethyl-5-phytylquinol
Tocopherol/tocotrienol Cyclase (TC)
delta-tocotrienol
gamma-Tocotrienol
gamma-Tocopherol
delta-Tocopherol
γ-TMT
Tocopherol/tocotrienol Gamma-methyltransferase (γ-TMT)
γ-TMT
beta-tocotrienol
alpha-Tocotrienol
alpha-Tocopherol
beta-Tocopherol

Figure 2.4 The biosynthetic pathway of tocopherol and tocotrienol with naturally occurring *α, β, γ,* and *δ* derivatives. This pathway was constructed by using the reference pathway "*Ubiquinone and other terpenoid quinone biosynthesis pathway*" on KEGG. The two substrates phytyl diphosphate (PDP) and homogentisate (HGA) are necessary for vitamin E biosynthesis. All enzymes are marked with red boxes: Hydroxyphenylpyruvate dioxygenase (*HPPD*), homogentisate prenyltransferase (*HPT*), 2-methyl-6-prenylbenzoquinol methyltransferase (*MSBQ*), tocopherol cyclase (*TC*) and tocopherol methyltransferase (*TMT*) are the major genes involved in the production of *α, β, γ,* and *δ* tocopherol.

The *δ* tocochromanols omits both methylation steps. The final methylation step occurs through the enzymes tocopherol/tocotrienol cyclase (*TC*) and tocopherol/tocotrienol methyltransferase to form *γ* and *α* tocochromanols (Hunter and Cahoon 2007). This process occurs on the third position of the benzoquinol ring to produce the final products of tocopherol and tocotrienols (Hunter and Cahoon 2007). The biosynthetic pathway of tocochromanols consists of several enzymes which are vital in the synthesis of tocochromanols, i.e. ρ-hydroxyphenylpyruvate dioxygenase (*HPPD*), homogentisate phytyltransferase (*HPT*), 2-methyl-6-phytyl-1,4-benzoquinone methyltransferase / 2-methyl-6-solanyl-1,4-benzoquinone methyltransferase (*MPBQ* / *MSBQ-MT*), tocopherol/tocotrienol cyclase (*TC*), and tocopherol/tocotrienols methyltransferase (*TMT*), but the enzymes *PK*, *GGR*, *HPPD* and *PrBQMT* are also used in other pathways (Hunter and Cahoon 2007). The biosynthetic enzymes of vitamin E are found within the chloroplast and chromoplast, but are also likely to be found in all other plastids (Hirschberg 1999, Ruhlman and Daniell 2007, Hunter and Cahoon 2007).

2.7 Genetics of vitamin E genes

2.7.1 Hydroxyphenylpyruvate dioxygenase (*HPPD*)

ρ-Hydroxyphenylpyruvate dioxygenase (*HPPD* aka *4-HPPD*) belongs to the family 2-oxoacid dependent dioxygenases (Ryle et al 2002, Lin et al 2013). It

is a mononuclear, non-heme, iron-containing enzyme which catalyses the conversion of 4-hydroxyphenylpyruvate (HPP) to 2,5-dihydroxyphenylacetate (HGA). The conversion step of HPP to HGA is an important step in the vitamin E biosynthetic pathway and was first recognized in *Arabidopsis thaliana* with the mutant *phytoene desaturation* (*PDS 1)* (Norris et al 1995, Nowicka and Kruk 2016). This key enzyme is involved in the synthesis of tocopherol and plastoquinone (Lin et al 2013). In plants the end product of the shikimate pathway (chorismate) is isomerized to prephenate, which leads to the production of the amino acid tyrosine (Hunter and Cahoon 2007). *HPPD* is the enzyme central to the biosynthesis of all quinoid compounds such as plastoquinones and tocopherols – both essential molecules in plants. Overexpression of *HPPD* in plants has a modest effect on increasing tocopherol levels in leaves and seeds (Rippert et al 2004). According to Chrost et al (1999) the expression levels of *HPPD* might be the factor controlling the tocopherol levels in plants, since mRNA levels of *HPPD* were increased as well as the tocopherol synthesis during senescence.

HPPD is known as a cytosolic gene and can also be useful as an herbicide (Gracia et al 1999, Siehl et al 2014). Studies show that redirecting *HPPD* to plastids does not enhance tocopherol content over the levels achieved by cytosolic expression (Hunter and Cahoon 2007). In another study conducted on soybean it was shown that HGA might be the limiting factor in vitamin E production, with the addition of exogenous HGA doubling the levels of tocopherol (Karunanandaa et al 2005). This suggests that elevating HGA concentrations could increase vitamin E biosynthesis, but it also indicates that *HPPD* could be a limiting factor in tocopherol biosynthesis.

2.7.2 Homogentisate phytyltransferase (*HPT*)

HPT catalyzes the committed step of tocopherol biosynthesis in the transferal of the prenyl group to HGA, which can either be performed by homogentisate phytyltransferase (*HPT*) or homogentisate geranylgeranyl transferase (HGGT), depending on the substrate. The prenyltransferase substrate determines the specific end product of tocopheromanols, i.e. whether

tocopherol or tocotrienols are produced by the biosynthetic pathway of vitamin E.

The prenytransferases are predicted integral membrane proteins that belong to the ubiA family (Hunter and Cahoon 2007). Indeed, *HPT* is a membrane bound chloroplast enzyme, which, making it a suitable candidate for an enzyme with a high flux coefficient (Collakova and DellaPenna 2003). Soll et al (1980) found activity within the inner membranes of both the chloroplast and chromoplasts when using cell fractionation methods.

The gene encoding *HPT1* and *HPT* has been cloned from *Synechocystis*, *Arabidopsis*, barley, wheat and rice. *HPT* has been transgenically expressed in plants, with the overexpression in *Arabidopsis* leaves causing an increase in the tocopherol content (Collakova and DellaPenna 2003). The overexpression of *HPT* was found to not only increase leaf tocopherol content, but also seed tocopherol content in *Arabidopsis* by up to 4.4-fold and the antisense suppression of *HPT* reduces the levels up to 10-fold in *Arabidopsis* seeds (Collakova and DellaPenna 2003). The overexpression of *HPT* and HGGT results in the enhancement of tocochromanols levels, thereby suggesting that *HPT* can assert flux control through the vitamin E biosynthetic pathway (Hunter and Cahoon 2007). The expression of barley HGGT in *Arabidopsis* leaves shifts the vitamin E biosynthesis from the production of tocopherols towards the tocotrienol production, increasing the total tocochromanol content (Cahoon et al 2003). This indicates that *HPT* is the key enzyme that can dictate whether tocopherol or tocotrienol are produced.

2.7.3 2-methyl-6-phytyl-1,4-benzoquinone methyltransferase / 2-methyl-6-solanyl-1,4-benzoquinone methyltransferase (*MPBQ / MSBQ-MT*)

The enzyme 2-methyl-6-phytyl-1,4-benzoquinone methyltransferase (aka 2-methyl-6-solanyl-1,4-benzoquinone methyltransferase, *MPBQ / MSBQ-MT*) is located in the inner membrane of the chloroplast. The enzymes for tocopherol synthesis are homologous in cyanobacteria and plants except for *MPBQ/MSBQ-MT* that catalyze a key methylation step in tocopherol and

plastoquinone (PQ) synthesis. The enzyme *MSBQ* catalyzes methylation of C-3 of 2-methyl-6-solanylbenzoquinol in the terminal step of PQ synthesis (Eitenmiller and Lee 2004).

Shintani et al (2002) identified this gene from the *Synechocystic* sp. PCC6803 genome that encodes methyltransferase. According to Hunter and Cahoon (2007) the *Synechocystic* gene was cloned via homology to the *Arabidopsis γ* tocopherol methyltransferase (*γ-TMT*) gene but only has 18% amino acid sequence similarity to the *Synechocystic* gene. The production of *γ* and *α* tocotrienols can be done by the addition of *HGGT* gene alone, but *MSBQ* appears to methylate either tocopherols or tocotrienols (Cahoon et al 2003). *MSBQ* produce *α* tocopherol in physiologically mature seeds of wildtype sunflower as well as *γ-TMT* activities in developing seeds (Shintani and DellaPenna 1998, Cheng et al 2003). The synthesis of *α* tocopherol is done by *MSBQ* and *γ-TMT*, but when both are mutated it disrupts the synthesis of *α* tocopherol and enhance *β, γ,* and *δ* tocopherol synthesis (Hass et al 2006). According to Cheng et al (2003) there are no *MPBQ/MSBQ-MT* mutations to redirect the flow in the biosynthetic pathway of tocopherol and modifying tocopherol composition through disrupted *PQ* biosynthesis. The overexpression of *MPBQ/MSBQ-MT* alters the composition of vitamin E and not the total content e.g. the overexpression in soybean seeds converts the pools *β* and *δ* tocopherols to *α*and *γ* tocopherol and not the total tocopherol content (Van Eenennaam et al 2003). Therefore, the enzyme *MPBQ/MSBQ-MT* was described as playing an important role in determining the tocopherol profile rather then determining the content.

2.7.4 *γ*-Tocopherol methyltransferase (*TMT*)

γ-Tocopherol methyltransferase (*TMT*) catalyzes the first step in the synthesis of *β* and *α* tocopherol. At the 5th position on the chromanol ring, the *γ* and *δ* tocopherol are methylated by a specific *γ-TMT* to yield *α* and *β* tocopherol. The methyltransferase activities of *α* and *β* have not been identified in nature and *α* and *β* tocopherol are considered to be terminal products of the biosynthesis.

Soll et al (1980) identified the methyltransferse activity of the gene *γ* tocopherol methyltransferase (*TMT*) in the 1980's that was located in the inner membrane of the spinach chloroplast. The isolation of the *TMT* gene was done using the *Synechocystis* homology to genes that are previously involved in the biosynthesis of vitamin E (Shintani and DellaPenna 1998, Cheng et al 2003). This homology was used to identify and clone the *Arabidopsis TMT* gene (Cheng et al 2003). The overexpression of *γ-TMT* in *Arabidopsis* does not affect the total tocopherol content but increases the *α* tocopherol content in dicot seeds. The expression of *E. coli* to *Arabidopsis TMT* methylates *δ* and *γ* tocopherol showing a three-fold preference for *γ* tocopherol but the overexpression of *Arabidopsis TMT* converts almost the entire content of *γ* tocopherol in *Arabidopsis* seeds to *α* tocopherol (Cheng et al 2003). Most dicots seeds like sunflower and soybean are rich in *γ* tocopherol, indicating that *TMT* activity are limited for *α* tocopherol synthesis in dicot seeds (Lu et al 2013). Thus understanding the role and activity of *TMT* explains why many oilseeds consist of low *α* tocopherol levels.

2.7.5 Tocopherol/tocotrienol cyclase (*TC*)

Tocopherol cyclase is located to the plastoglobules in the chloroplast and is the second enzyme in the committed vitamin E biosynthesis pathway. This enzyme is very important in the formation of the chromanol ring structure, generating an additional oxygen heterocycle for the prevention of oxidation. The chromanol head group forms a benzoquinol intermediate from the enzyme *TC* that catalyzes the formation of *δ* tocopherol from 2-methyl-6-phytylbenzoquinone and *γ* tocopherol from 2,3-dimethyl-6-phytylbenzoquinone (Shintani et al 2002). *TC* recognizes the OH group at C1 of the hydroquinone, the E-configuration of the double bond and the length of the lipophilic side chain on 2-methyl-6-phytylbenzoquinone / 2,3-dimethyl-6-phytylbenzoquinonol as the three main substrate features. The enzyme *TC* is effective in converting 2,3-dimethyl-6- geranylbenzoquinonol to *γ* tocopherol 3 (*γ*-T3) and 2-methyl-6-geranylgeranyl benzoquinone to *δ* tocopherol 3 (*δ*-T3). The investigation done by Koch et al (2002) indicated that the methylation

capacity of *δ* and *γ*-T3 is almost equivalent to the capacity to methylate the corresponding tocopherols.

According to Sattler et al (2003) both *δ* and *γ* tocopherol are found in plants and they are presumed to cyclize mono- and di-methylated prenylbenzoquinols. Genes for *TC* have been cloned and expressed from *Synechocystis* sp *PCC6803*, *Arabidopsis*, *Zea mays,* as well as cyanobacteria (Stocker et al 1996, Sattler et al 2003). The *TC* gene and prenyltransferases plays an important role in flux control through the pathway (Hunter and Cahoon 2007). This is indicated through the total leaf tocopherol content that increases seven-fold when overexpressed in *Arabidopsis* (Hunter and Cahoon 2007).

2.8 Sunflower tocopherol genes

Kamal-Eldin and Andersson (1997) found that oil with high linolenic acid (18:3) has low *α* tocopherol levels and that there is a positive correlation between linoleic acid (18:2) and *α* tocopherol. However, there are two genes responsible for the tocopherol composition in sunflower oil, one controlling the *α, β* tocopherol and the other the *γ and δ* tocopherol (Demurin 1993). Vera-Ruiz et al (2006) identified a sunflower line T589 that is characterized by an increase of *β* tocopherol content, which is determined by the gene Tph1. This gene mutant was found on linkage group 1 of the sunflower genetic map. The Tph1 gene is believed to show more balanced *in vitro* and *in vivo* antioxidant properties and also control *α* and *β* tocopherol.

Tang et al (2006) found that the *Tph1* gene mutant partially disrupts the synthesis of *α* tocopherol in sunflower. According to Cheng et al (2003) the synthesis of *α* and *γ* tocopherol can be disrupted without impairing plant growth and development. The *Tph1* gene mutant is also believed to interfere in the methyltransferase activity that is needed for the synthesis of *α* and *γ* tocopherol. Further research on the *Tph1* mutant led to the isolation of two 2-methyl-6-phytol-1,4-benzoquinone / 2-methyl-6-solanyl-1,4-benzoquinone methyltransferase (*MPBQ/MSBQ-MT*) paralogs from sunflower (Tang et al

2006). They also identified *Tph1* as a non-lethal knockout mutant of *MT1* that's caused by the insertion of a 5.2Kb Ty3/gypsy-like-retro transposon on exon 1 which caused a significant increase in *β* tocopherol in mutant sunflower inbred lines.

Škorić et al (2007) generated sunflower genotypes with high oleic hybrids and altered tocopherol profiles by incorporating the *ol* + *Tph1* genes. They modified these genotypes to breed for high seed and oil yields. Škorić et al (2007) found that oil containing these hybrids had much longer shelf life than standard sunflower oil. A sunflower germplasm with high *γ* tocopherol was developed through a single locus *Tph2* (Garcia-Moreno et al 2012). This gene mutant is believed to be responsible for controlling the *γ* and *δ* tocopherol content. Garcia-Moreno et al (2012) found that modifying the *γ-TMT Tph2* gene would lead to unstable expression of high *γ* tocopherol content.

Haddadi et al (2011) used *A. thaliana* sequences to find sunflower homologues to identify any genomic sequences for vitamin E biosynthesis. They found some sunflower sequences which could directly or indirectly play a role in tocopherol biosynthesis. Five QTL's were found on linkage groups 1, 8, 10 and 14 for total tocopherol content that accounted for 45% of the phenotypic variation (Haddadi et al 2011). The four genes found namely, *VTE4*, *HPPD*, GST and Doug1 exhibited co-localization with the QTL for total tocopherol content. According to Haddadi et al (2011) the genes *HPPD*, *VTE4* and *SMT2* could be used for altering phytosterol and tocopherol content in sunflower seeds by developing functional markers.

2.9 Sunflower genomics

Efforts to obtain the complete sequence of the sunflower genome are underway (Kane et al 2010). The sunflower genome is estimated to be 3.5 billion bp in size – a little larger than the human genome of 3.2 GB (Arumuganathan and Earle 1991, Baack et al 2005, Kane et al 2011). According to Kane et al (2011), sequencing the sunflower genome will help in developing a reference genome that can be used to study the 24 000 species

within the sunflower family (Asteraceae). Obtaining the sunflower genome will also be very important for possible improvements in sunflower breeding programs aiming to improve the quality of hybrids and improving resistance to diseases and other stress factors, as well as yield and oil content (National sunflower association, http://www.sunflowernsa.com). Sunflower has a diploid genome, with a base chromosome number of 17 (Lu et al 2007, Kane 2010).

GenBank (http://www.ncbi.nlm.nih.gov/nucest/?term=Sunflower) contains more than 40 000 sunflower ESTs that have been assembled from multiple libraries using the CAP3 program (Kozik et al 2003). This program was organized into the "*Composite Genome Project*" database that represents an assembly of more than 12 000 sunflower unigenes. Various other studies found ESTs to identify embryogenesis-related genes (Tamborindeguy et al 2004) and organ-specific ESTs (Fernández et al 2003). These ESTs and other genomic approaches will promote the process of developing new sunflower trait genes, improving molecular breeding and the genetic engineering of sunflower (Lu et al 2007).

Successful incorporation of genetic traits into sunflower will rely on understanding the genome of this crop through transcriptome studies, modern breeding programs by using molecular markers generated via random fragment length polymorphism (RFLP), marker assisted selection (MAS), simple sequence repeats (SSRs), random amplified polymorphic DNA (RAPD) and by single nucleotide polymorphism (SNP) techniques (Tanksley et al 1989, Rommens and Kishore 2000). These breeding tools and other genomic approaches could lead to the discovery of genes with the potential of improving phenotypic (economic) traits (Sung et al 2003, Kane et al 2011).

2.10 Conclusion

Sunflower oil is regarded as one of the most important oil seed crops in South Africa (van der Merwe 2013), with oil content and degree of fatty acids determing the oil's quality (Kilcast and Subramaniam 2000). Its oil content is characterized by high concentrations of linoleic and oleic acid, which

contributes ± 90% of the oil composition. Sunflower oil also consists of saturated fatty acid e.g. stearic and palmitic acid that contributes 10% to the total fatty acid composition. These high levels of unsaturated fatty acid (± 90%) makes sunflower oil highly susceptible to oxidation, resulting in oil rancidity, off flavour and unstable oil. Oxidation can be retarded or prevented by the precense of vitamin E, which displays of distinct antioxidant property (Herrera and Barbas 2001).

Vitamin E consists of various tocopherol (*α, β, γ* and *δ*) and tocotrienol (*α, β, γ* and *δ*) derivatives (Hofius and Sonnewald 2003). Tocopherol and tocotrienol is collectively known as tocochromanols, which share the same biosynthetic pathway, derived from the shikimate pathway (Hofius and Sonnewald 2003). The pathway consists of various enzymes (e.g *HPPD, HPT, MSBQ, TC* and *TMT*) that lead to the development of the four tocopherol derivatives. These derivatives vary between individual plants. Rippent et al (2004) found that the overexpression of *HPPD* causes an increase of tocopherol levels in seeds and leaves. *HPT* catalyses the committed steps of tocopherol synthesis - depending on the subtrate used (e.g *HPT* or *HGGT*). Collakova and DellaPenna (2003) found that the overexpression of *HPT* results in an increase in leave and seed tocopherol content in *Arabidopsis.* According to Van Ennennaam the gene *MSBQ* is important for determining tocopherol composition and not tocopherol content. *TMT* is important for *α* tocopherol synthesis in dicot seeds but when overexpressed it increases *α* tocopherol content in seeds (Cheng et al 2003). The gene *TC* methlates *δ* and *γ* tocopherol-3 and plays important roles in flux control. Literature indicates that the sunflower genes *Tph1* and *Tph2* are involved in tocopherol composition where *Tph1* controls *α* and *β* tocopherol and *Tph2* controls *γ* and *δ* tocopherol. The availability of the sunflower genome has the potential to improve important economical traits e.g. oil quality, yield and other factors i.e disseases. The incorporation of existing mutantions, e.g. *Tph1* and *Tph2*, and new mutations in sunflower hybrids will improve and increase sunflower oil quality (e.g. oxidative stability and shelf life) (Kane et al 2011).

References

Abdalla AE and Roozen JP 1999. Effect of plant extracts on the oxidative stability of sunflower oil and emulsion. *Food Chemistry* 323-329

Abitogun AS and Oshodi AA 2010. Effects of degumming and bleaching on physicochemial properties of crude sunflower oil. *Electronic Journal of Environmental, Agricultural and Food Chemistry* 1145-1151

Aggarwal BB, Sundaram C, Prasad S and Kannappan R 2010. Tocotrienols, the vitamin E of the 21st century: Its potential against cancer and other chronic diseases. *Biochemical Pharmacology* 1613-1631

Aluyor EO and Ori-Jesu M 2008. The use of antioxidants in vegetable oils – A review. *African Journal of Biotechnology* 4837-4841

Arumuganathan K and Earle ED 1991. Nuclear DNA content of some important plant species. *Plant Molecular Biology* Reporter 208-218

Awatif I and Shaker M 2014. Quality characteristics of high-oleic sunflower oil extracted from some hybrids cultivated under Egyptian conditions. *Helia* 113-126

Azizkhani M, Kamkar A and Mozaffari Nejad AS 2011. Effects of tocopherols on oxidative stability of margarine. *Journal of the Chemical Society of Pakistan* 134-137

Baack EJ, Whitney KD and Rieseberg LH 2005. Hybridization and genome size evolution: Timing and magnitude of nuclear DNA content increases in *Helianthus* homoploid hybrid species. *New Phytologist* 623-630

Babovic N, Žižovic I, Saičic S, Ivanovic J, and Petrovic S 2010. Oxidative stabilization of sunflower oil by antioxidant fractions from selected *Lamiaceae* herbs. *Chemical Industry and Chemical Engineering Quarterly* 287-293

Baldwin WS, Minneapolis, and Keeney KW 1974. Methylation of tocopherols. *United States Patent (3819657)* 1-2

Bandarra NM, Campos RM, Batista I, Nunes ML and Empis JM 1999. Antioxidant synergy of *α* tocopherol and phospholipids. *Journal of the American Oil Chemists Society* 905-913

Berglund D 2007. Sunflower production. NDSU Extension service N.D *Agricultural experiment station*, North Dakota State University 1-117

Boscoboinik D, Szewczyk A and Azzi A 1991. *α* -Tocopherol (Vitamin E) regulates vascular smooth muscle cell proliferation and protein kinase C activity. *Archives of Biochemistry and Biophysics* 264-269

Burton GW and Ingold KU 1981. Autoxidation of Biological Molecules 1. The antioxidant activity of vitamin E and related chain-breaking phenolic antioxidants *in vitro*. *Journal of the American Oil Chemists Society* 6472-6477

Cahoon EB, Hall SE, Ripp KG, Ganzke TS, Hitz WD and Coughlan SJ 2003. Metabolic redesign of vitamin E biosynthesis in plants for tocotrienol production and increased antioxidant content. *Nature Biotechnology* 1082-1087

Codex Alimentarius Commission 2001. Codex Standard for margarine 2nd edition, *Codex standard* 32 1-4

Collakova E and DellaPenna D 2003. Homogentisate phytyltransferase activity is limiting for tocopherol biosynthesis in *Arabidopsis*. *Plant Physiology* 632-642

Colombo ML 2010. An update on vitamin E, tocopherol and tocotrienol-perspectives. *Molecules* 2103-2113

Corsini MS, Silva MG and Jorge N 2009. Loss in tocopherols and oxidative stability during the frying of frozen cassava chips. *International Journal of Fats and Oils* 77-81

Cheng Z, Sattler S, Maeda H, Sakuragi Y, Bryant DA, DellaPenna D 2003. Highly divergent methyltransferases catalyze a conserved reaction in tocopherol and plastoquinone synthesis in cyanobacteria and photosynthetic eukaryotes. *The Plant Cell* 2343-2356

Chrost B, Falk J, Kernebeck B, Mölleken H and Krupinska K 1999. Tocopherol biosynthesis in senescing chloroplasts: A mechanism to protect envelope membranes against oxidative stress and a prerequisite for lipid remobilization? In: *Argyroudi-Akoyunoglou JH, Senger H, editors. The chloroplast*: *Molecular Biology to Biotechnology* 171-176

Ciemniewska-Żytkiewicz H, Ratusz K, Bryś J, Reder M and Koczoń P 2014. Determination of the oxidative stability of hazelnut oils by PDSC and Rancimat methods. *Journal of Thermal Analysis and Calorimetry* 875-881

Department of Agriculture Forestry and Fishery 2010. Sunflower seed. Market value chain profile. Department of agriculture 204-230

Department of Agriculture Forestry and Fishery 2013. Quarterly sunflower market analysis and outlook, bulletin 3 of 2013. (http://www.nda.agric.za/doaDev/sideMenu/Marketing/Quartely%20Bulletins%20Uploaded/Sunflower%20Bulletin%20No.3%20of%202013.pdf)

Department of Agriculture Forestry and Fishery 2015. Weekly price watch, directorate: Statistics and economic analysis.

(http://www.daff.gov.za/Daffweb3/Portals/0/Price%20Watch/PriceWatch%202015-05-29.pdf)

Dedio W 1985. Effects of seeding and harvesting dates on yield and oil quality of sunflower cultivars. *Canadian Journal of Plant Science* 299–305

Demurin Y 1993. Genetic variability of tocopherol composition in sunflower seed. *Helia* 59-62

Eitenmiller R and Lee J 2004. Vitamin E. Food chemistry, composition and analysis. *Marcel Dekker Incorporation, Food Science and Technology* 1-32

Epstein SS, Forsyth J, Saporoschetz IB and Mantel N 1966. An exploratory investigation on the inhibition of selected photosensitizers by agents of varying antioxidant activity. *Radiation Research* 322-335

Evans HM and Bishop KS 1922. Fetal resorption. *Science* 650

Falk J and Munne´-Bosc S 2010. Tocochromanol functions in plants: antioxidation and beyond. *Journal of Experimental Bot*any 1549-1566

Fernández-Martínez J, Jimenez A, Dominguez J, Garcia JM, Garces R, Mancha M 1989. Genetic analysis of the high oleic acid content in cultivated sunflower (*Helianthus annuus L.*). *Euphytica* 39-51

Fernandez-Martinez JM, Perez-Vich B and Velasco L 2009. Mutation breeding for oil quality improvement in sunflower. Induced Plant Mutations in the Genomics Era. *Plant Breeding Section, Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture, International Atomic Energy Agency* 177-181

Fick GN 1978. Selection for self-fertility and oil percentage in development of sunflower hybrids. In: *Proceedings of the 8th International Sunflower Conference*. *International Sunflower Association*, *Paris* 418-422

Fick GN 1984. Inheritance of high oleic acid in the seed oil of sunflower. *In proceedings of the 6th International Sunflower Forum*. *National Sunflower Association*, *Bismarck North Dakota* 9

Frega N, Mozzon M and Lercker G 1999. Effects of Free Fatty Acids on Oxidative Stability of Vegetable Oil. *Journal of American Oil Chemists' Society* 325-329

Friedt W, Ganssmann M and Korell M 1994. Improvement of sunflower oil quality. *In Proceeding of EUCARPIA Symposium on Breeding of Oil and Protein Crops*, *Albena, Bulgaria* 1-30

Fuller M, Diamond J and Apple WT 1967. High oleic sunflower oil, stability and chemical modification. *Journal of American Oil Chemists' Society* 264-267

Fuster M, Lampi A, Hopia A and Kamal-Eldin A 1998. Effects of α and γ tocopherols on the auto-oxidation of purified sunflower triacylglycerols. *Lipids* 715-722

Garcia I, Rodgers M, Pepin R, Hsieh TF and Matringe M 1999. Characterization and subcellular compartmentation of recombinant 4-hydroxyphenylpyruvate dioxygenase from *Arabidopsis* in transgenic tobacco. *Plant Physiology* 1507-1516

Goffner D, Cazalis R, Percie du Sert C, Calmes J and Cavalie G 1988. 14C photassimilate partitioning in developing sunflower seeds. *Journal of Experimental Bot*any 1411-1420

Goffman FD and Bohme T 2001. Relationship between fatty acid profile and vitamin E content in maize hybrids (*Zea mays* L). *Journal of Agricultural and Food Chemistry* 4990-4994

Gunstone E 2002. Sunflower seed and its products. *Information* 159-163

Hamilton RJ, Kula C, McNeill GP, Padley FB and Pierce JH 1998. Effects of tocopherols, ascorbyl palmitate, and lecithin on autoxidation of fish oil. *Journal of American Oil Chemists' Society* 813-823

Harris HC, McWilliam JR and Mason WK 1978. Influence of temperature on oil content and composition of sunflower seed. *Australian Journal of Agricultural Research* 1203-1212

Hass CG, Tang S, Leonard S, Traber MG, Miller JF and Knapp SJ 2006. Three non-allelic epistatically interacting methyltransferase mutations produce novel tocopherol (vitamin E) profiles in sunflower. *Theoretical and Applied Genetics* 767-782

Heiser CB Jr. 1978. Taxonomy of Helianthus and origin of domesticated sunflower. In: Sunflower science and technology. *American Society of Agronomy, Crop Science Society of America, Soil Science Society of America* 31-53

Herting DC and Drury EE 1963. Vitamin E content of vegetable oils and fats. *Journal of Nutrition* 335-342

Herrera E and Barbas C 2001. Vitamin E: Action, metabolism and perspectives. *Journal of Physiology and Biochemistry* 43-56

Hess JL 1993. Vitamin E, α tocopherol. In: *Alscher RG and Hess JL edition. Antioxidants in higher plants. Boca Raton, FL: CRC Press* 111-134

Hewezi T, Le´ger M, Kayal W and Gentzbittel L 2006. Transcriptional profiling of sunflower plants growing under low temperatures reveals an extensive down-regulation of gene expression associated with chilling sensitivity. *Journal of Experimental Bot*any 3109-3110

Hofius D and Sonnewald U 2003. Vitamin E biosynthesis: Biochemistry meets cell biology. *Trends in Plant Science* 6-8

Hirschberg J 1999. Production of high-value compounds: Carotenoids and vitamin E. *Biotechnology* 186-191

Hunter SC and Cahoon EB 2007. Enhancing vitamin E in oilseeds: Unravelling tocopherol and tocotrienol biosynthesis. *Lipids* 97-108

Ikeda N and Fukuzumi K 1977. Synergistic antioxidant effect of nucleic acids and tocopherols. *Journal of American Oil Chemists' Society* 360-366

Iuoras M, Vrânceanu AV, Stanciu D and Sorega I 2002. Sunflower inbred lines derived from interspecific hybrids. *Helia* 59-64

Janick J 2013. Development of new world crops by indigenous Americans. *American Society of Horticultural Science* 406-412

Jorge N, Márquez-Ruiz G, Martín-Polvillo M, Ruiz- Méndez MV and Dobarganes MC. 1996a. Influence of dimethylpolysiloxane addition to edible oils: Dependence on the main variables of the frying process. *International Journal of Fats and Oils* 14-19

Jung MY and Min DB 1990. Effects of α, γ and δ tocopherols on oxidative stability of soybean oil. *Journal of Food Science* 1464-1465

Kabel AM 2014. Free radicals and antioxidants: Role of enzymes and nutrition. *World Journal of Nutrition and Health* 35-38

Kamal-Eldin A and Appleqvist LA 1996. The chemistry and antioxidant properties of tocopherols and tocotrienols. *Lipids* 1-31

Kamal-Eldin A and Andersson R 1997. A multivariate study of the correlation between tocopherol content and fatty acid composition in vegetable oils. *Journal of American Oil Chemists' Society* 375-380

Kane NC 2010. Genome BC. Sunflower genome holds the promise of sustainable agriculture. *ScienceDaily* www.sciencedaily.com/releases/2010/01/100112121930.htm

Kane NC, Gill N, King MG and Bowers JE 2011. Progress towards a reference genome for sunflower. *Botany* 429-437

Karlovic` D, Recseg K, Kovari K, Nobik-Kovacs A, Skoric` D and Demurin Y 1997. Characteristic quality and oxidative stability of sunflower oil with altered fatty acid and tocopherol composition. *22nd World Congress and Exhibition of International Society for Fat Research*, Kuala Lumpur, Malaysia

Karunanandaa B, Qi Q, Hao M, Baszis SR, Jensen PK, Wong YH, Jiang J, Venkatramesh M, Gruys KJ, Moshiri F, Post-Beittenmiller D, Weiss JD and Valentin HE 2005. Metabolically engineered oilseed crops with enhanced seed tocopherol. *Metabolic Engineering* 384-400

Khan S, Lisa SA, Obaid M and Chowdhury K 2015. Tocopherol content of vegetable oils/fats and their oxidative deterioration during storage. *World Journal of Pharmacy and Pharmaceutical Sciences* 1537-1548

Kilcast D and Subramaniam P 2000. The stability and shelf life of food. *Woodhead Publishing Limited* 1-19

Killi F and Altunbay SG 2005. Seed yield, oil content and yield components of confection and oilseed sunflower (*Helianthus annuus* L.) cultivars planted in different dates. *International Journal of Agriculture and Biology* 1-1

Koch M, Arango Y, Mock HP and Heise KP 2002. Factors influencing α tocopherol synthesis in pepper fruits. *Plant Physiology* 159: 1015-1019

Kozik A, Chan B and Michelmore R 2003. Lettuce / Sunflower EST CGPDB project: DNA analysis, assembly visualization and validation. *Plant Animal Genomes Conference* 886

Lacombe S and Berville A 2001. A dominant mutation for high oleic acid content in sunflower (*Helianthus annuus* L.) seed oil is genetically linked to a single oleate-desaturase RFLP locus. *Molecular Breeding* 129-137

Lea C and Ward R 1959. Relative antioxidant activities of seven tocopherols. *Journal of The Science of Food and Agriculture* 537-548

Leeuwner DV 2005. Genotype x environment interaction for sunflower hybrids in South Africa. 5-11

Liebler DC 1993. The role of metabolism in the antioxidant function of vitamin E. *Critical Reviews in Toxicology* 147-169

Lin J, Sheih Y, Chang T, Chang N, Chang C, Shen C and Lee H 2013. The interactions in the carboxyl terminus of human 4-hydroxyphenylpyruvate dioxygenase are critical to mediate the conformation of the final helix and the tail to shield the active site for catalysis. *PloS one* 1-9

Lobo V, Patil A, Phatak A and Chandra N 2010. Free radicals, antioxidants and functional foods: Impact on human health. *Pharmacognosy reviews* 118-126

Lu G, Hu X and Bidney D 2007. Sunflower Chapter 1.3 In: *Transgenic Crops VI: Biotechnology in Agriculture and Forestry* 39-51

Lu Y, Rijzaani H, Karcher D, Ruf S and Bock R 2013. Efficient metabolic pathway engineering in transgenic tobacco and tomato plastids with synthetic multigene operons. *Proceedings of the National Academy of Sciences of the United States of America* 623-632

Mackenzie CG, Mackenzie JB and McCollum EV 1939. Growth and reproduction on a low fat diet. *Biochemical Journal* 935-943

Márquez-Ruiz G, Martin-Polvillo M and Dobarganes C 2003. Effect of temperature and addition of α tocopherol on the oxidation of trilinolein model systems. *Lipids* 233-240

McLaren ME 2014. The values of working in the Alberta oil sands. Author house 103-104

Méndez E, Sanhueza J, Speisky H and Valenzuela A 1996. Validation of the Rancimat test for the assessment of the relative stability of fish oils. *Journal of American Oil Chemists' Society* 1033

Mermelstein NH 2010. Improving soybean oil. *Food safety and quality* 72-76

Miller BC, Oplinger ES, Rand R, Peters J and Weis G 1984. Effects of planting date and planting population on sunflower performance. *Agronomy Journal* 511-515

Munné-Bosch S, Weiler EW, Alegre L, Müller M, Düchting P and Falk J 2007. *α* -Tocopherol may influence cellular signalling by modulating jasmonic acid levels in plants. *Planta An international Journal of Plant Biology* 681-691

National Agricultural Marketing Council. Market and economic research centre, Food price monitor 2016.
http://www.namc.co.za/upload/food_price_monitoring/NAMC Food%20Price%20Monitor-Feb%202016.pdf

National sunflower association. Sunflower oil http://www.sunflowernsa.com/oil/

Norris SR, Barrette TR and DellaPenna D 1995. Genetic dissection of carotenoid synthesis in *Arabidopsis* defines plastoquinone as an essential component of phytoene desaturation. *The Plant Cell* 2139-2149

Nowicka B and Kruk J 2016. Cyanobacteria use both p-hydroxybenozate and homogentisate as a precursor of plastoquinone head group. *Proceedings of the National Academy of Sciences of the United States of America* 623-632

O' Brien RD 2004. Fats and oils. In: *Formulation and processing for application 2nd edition*. *CRC Press, London and New York* 235-474

O'Connor T and O'Brien N 2006. Lipid oxidation. In: *Advanced dairy chemistry volume 2, third edition, Lipids*. Edited by Fox PF and McSweeney PLH, *Springer, New York* 557-600

Putt E 1978. History and present world status. Chapter 1: In *Sunflower science and technology* 1-25

Raiola, A, Tenore GC, Barone A, Frusciante L and Rigano MM 2015. Vitamin E content and composition in tomato fruits: Beneficial roles and bio-fortification. *International Journal of Molecular Sciences* 29250-29264

Rengaraj D and Hong YH 2015. Effects of dietary vitamin E on fertility functions in poultry species. *International Journal of Molecular Sciences* 9910-9921

Rimbach G, Minihane AM, Majewicz J, Fischer A, Pallauf J, Virgli F and Weinberg PD 2002. Regulation of cell signalling by vitamin E. *Proceedings of the Nutrition Society* 415-425

Rippert P, Scimemi C, Dubald M and Matringe M 2004. Engineering plant shikimate pathway for production of tocotrienol and improving herbicide resistance. *Plant Physiology* 92-100

Rizvi S, Raza ST, Ahmed F 2014. The role of vitamin E in human health and some diseases. *Sultan Qaboos University Medical Journal* 157-165

Robbins K and Sewalt V 2005. Extending freshness with rosemary extract. *Food Technology (Inform)* 534-535

Rondanini D, Savin R and Hall AJ 2003 Dynamics of fruit growth and oil quality of sunflower (*Helianthus annuus* L.) exposed to brief intervals of high temperature during grain filling. *Field Crop Research* 79-90

Romero N, Robert P, Masson L, Ortiz J, Pavez J, Garrido C, Foster M, and Dobarganes C 2004. Effect of α tocopherol and α tocotrienol on the performance of Chilean hazelnut oil (*Gevuina avellana* Mol) at high temperature. *Journal of the Science of Food and Agriculture* 943

Rommens CM and Kishore GM 2000. Exploiting the full potential of disease resistance genes for agricultural use. *Current Opinion in Biotechnology* 120-125

Ruhlman T and Daniell H 2007. Plastid pathways (metabolic engineering via the chloroplast genome). *Applications of Plant Metabolic Engineering* 79-108

Ryle MJ and Hausinger RP 2002. Non-heme iron oxygenases. *Current Opinion in Chemical Biology* 193-201

Sadoudi R, Ammouche A, Ali AD 2014. Thermal oxidative alteration of sunflower oil. *African Journal of Food Science* 116-121

Sattler SE, Cahoon EB, Coughlan SJ and DellaPenna D 2003. Characterization of tocopherol cyclases from higher plants and cyanobacteria. Evolutionary implications for tocopherol synthesis and function. *Plant Physiology* 2184-2195

Sattler SE, Gilliland LU, Magallanes-Lundback M, Pollard M, and DellaPenna D 2004. Vitamin E is essential for seed longevity and for preventing lipid peroxidation during germination. *The Plant Cell* 1419-1432

Sattler SE, Me`ne-Saffrane´ L, Farmer EE, Krischke M, Mueller MJ, and DellaPenna D 2006. Nonenzymatic lipid peroxidation reprograms gene expression and activates defence markers in *Arabidopsis* tocopherol deficient mutants. *The Plant Cell* 3706-3720

Scharp WR 1986. Opportunities for biotechnology in the development of new edible vegetable oil products. *Journal of American oil chemist's society* 594-600

Schnable PS and Wise RP 1998. Male sterility and fertility restoration. *Trends in plant science* 175-180

Schultz G, Heintze A, Hoppe P, Hagelstein P, Go¨ rlach J, Meereis K, Schwanke U and Preiss M 1991. Tocopherol and carotenoid synthesis in chloroplasts: Tight linkage to plastidic carbon metabolism in developing chloroplasts. In: *Pell E, Steffen K, edition. Active oxygen/oxidative stress and plant metabolism*. Rockville, MD: *American Society of Plant Physiologists* 156-170

Shahidi F, Ho CT and Van Chuyen N 2013. Process-induced chemical changes in food. *Advances in Experimental Medicine and Biology* 1-372

Shintani and DellaPenna 1998. Elevating the vitamin E content of plants through metabolic engineering. *Science* 2098-2100

Shintani DK, Cheng Z and DellaPenna D 2002. The role of 2- methyl-6-phytylbenzoquinone methyltransferase in determining tocopherol composition in *Synechocystis* sp. PCC6803. *Federation of European Biochemical Societies Letters* 1-5

Siehl DL, Tao Y, Albert H, Dong Y, Heckert M, Madrigal A, Lincoln-Cabatu B, Lu J, Fenwick T, Bermudez E, Sandoval M, Horn C, Green JM, Hale T, Pagano P, Clark J, Udranszky IA, Rizzo N, Bourett T, Howard RJ, Johnson DH, Vogt M, Akinsola G and Castle LA 2014. Broad 4-hydroxyphenylpyruvate dioxygenase inhibitor herbicide tolerance in soybean with an optimized enzyme and expression cassette. *Plant Physiology* 1162-1176

Škorić D, Jocić S, Hladni N and Vannozzi GP 2007. An analysis of heterotic potential for agronomically important traits in sunflower (*Helianthus annuus L.*) *Helia* 55-74

Škorić D, Jocic S, Sakac Z and Lecic N 2008. Genetic possibilities for altering sunflower oil quality to obtain novel oils. *Canadian Journal of Physiology and Pharmacology* 215-220

Soll J, Kemmerling M and Schultz G 1980. Tocopherol and plastoquinone synthesis in spinach chloroplasts subfractions. *Archives of Biochemistry and Biophysics* 544-550

Stevens PF 2001 onwards. Angiosperm Phylogeny Website. Version 12, July 2012 (http://www.mobot.org/MOBOT/research/APweb/)

Stocker A, Fretz H, Frick H, Ru¨ ttimann A and Woggon WD 1996. The substrate specificity of tocopherol cyclase. *Bioorganic and Medicinal Chemistry* 1129-1134

Sung D-Y, Kaplan F, Lee KJ and Guy CL 2003. Acquired tolerance to temperature extremes. *Trends in Plant Science* 179-185

Tanksley SD, Young ND, Paterson AH and Bonierbale MW 1989. RFLP mapping in plant breeding: New tools for an old science. *Nature Biotechnology* 257-264

Tavassalkar S, Mishra H and Madhavan S 2012. Evaluation of antioxidant efficacy of natural plant extracts against synthetic antioxidants in sunflower oil. *Journal of Food Processing and Technology* 1-5

van der Merwe R, Labuschagne MT, Herselman L and Hugo A 2013. Stability of seed oil quality traits in high and mid-oleic acid sunflower hybrids. *Euphytica* 1-12

Van Eenennaam AL, Lincoln K, Durrett TP, Valentin HE, Shewmaker CK, Thorne GM, Jiang J, Baszis SR, Levering CK, Aasen ED, Hao M, Stein JC, Norris SR and Last RL 2003. Engineering vitamin E content: From *Arabidopsis* mutant to soy oil. *Plant Cell* 3007-3019

Vranceanu V and Stoenescu F 1971. Pollen fertility restorer gene from cultivated sunflower (*Helianthus annuus L.*). *Euphytica* 536-541

Warner K, Orr P and Glynn M 1997. Effect of fatty acid composition of oils on flavor and stability of fried foods. *Journal of the American Oil Chemists' Society* 347-356

Zandi P and Gordon MH 1999. Antioxidant activity of extracts from old tea leaves. *Food Chemistry* 285-288

Zielińska A and Nowak I 2014. Tocopherols and tocotrienols as vitamin E. *Science* 589-591

Chapter 3

Determining the correlation between tocopherols and unsaturated fatty acids in selected sunflower accessions

3.1 Abstract

Linchay Daniels [1], **Shaun Reeksting**[2] **& Dirk (ZH) Swanevelder**[1,3]
[1] Agricultural Research Council, Biotechnology Platform, Onderstepoort, Pretoria, South Africa
[2] Agricultural Research Council, Residue Laboratory, Onderstepoort Veterinary Institute, Pretoria, South Africa
[3] Research Fellow, College of Agriculture and Environmental Sciences, UNISA, Pretoria, South Africa

Email: Swanevelderd@arc.agric.za

Sunflower oil contains high levels of polyunsaturated fatty acids that increase its susceptibility to oxidation, which leads to rancid oil formation. Oxidation instability can be prevented by the presence of tocopherols, i.e. Vitamin E, in the oil. The aim of this chapter is to determine the correlation between the different tocopherols and unsaturated fatty acids (oleic and linoleic) of 104 sunflower accessions. Of all four tocopherol derivatives tested, *α* tocopherol was predominant with a correlated value of $r = 0.89$ with total tocopherols and a mean average value of 331.41 ppm across all lines. An increase in the tocopherol content observed, coincided with a decrease in oleic and linoleic acid levels, and *vice versa*. No correlation was observed between linoleic acid, *α, β, γ* and *δ* tocopherol. A positive correlation was found between oleic acid, *γ* (r=0.17) and *δ* (r=0.23) tocopherol but none between oleic acid, *α* and *β* tocopherol.

Key words: sunflower, oxidation, vitamin E, oleic acid, linoleic acid, *α, β, γ* and *δ* tocopherol, correlation

3.2 Introduction

Sunflower is the fourth most important oil crop in the world after soybean, palm oil and canola (USDA 2012). This crop grows well in virtually all summer rainfall areas of South Africa and is well adapted in both hot and dry conditions (Pannar seed 2006). In South Africa, sunflower is the third largest grain crop produced after maize and wheat (Department of Agriculture Forestry and Fishery 2011). The gross value for sunflower seed production amounted to R163 million in South Africa over the past 6 years (2008-2014) (NDA 2015). This crop is produced in eight of the nine South African provinces, with seeds primarily being used for sunflower oil pressing (Department of Agriculture Forestry and Fishery 2011). High quality sunflower oil is destined for human consumption, with the pressed oilcake used for manufacturing of animal feed such as oilcake meal (Department of Agriculture Forestry and Fishery 2011).

Parameters that play an important role in the lipid nutritional value are the fatty acid composition of the oil, the distribution pattern of the fatty acids (e.g. stereochemical position of the fatty acids in the three positions of the triacylglycerol molecule), and the total content and composition of tocopherols (Fernandez-Martinez et al 2009). Nutritionally, both oleic and linoleic acid are essential fatty acids, but oleic acid is preferred due to its high antioxidant properties (Fernandez-Martinez et al 2009). Fats, oils or any lipid containing foods for that matter, are easily oxidized by air contact, though at different rates, thereby contributing to the reduction of shelf life (Frega et al 1999). Lipid oxidation does not only cause unpleasant smells and flavour but also cause degradation of products that can result in harmful side effects on human health (Ying zhang et al 2010). These toxic compounds can cause health problems, such as aging, cancer and heart disease (Ammari et al 2012). The addition of antioxidants however, may retard or even prevent these health problems (Cosgrove et al 1987). This is achieved as a result of the effect antioxidants have on the oxidation process. Antioxidants are substances that “substantially prevent or delays oxidation”, even when present in low concentrations (Hallliwell et al 1995). Oxidation in organic

systems produce free-radicals that start a chain reaction which attack healthy cells resulting in loss of structure and function (Aluyor and Ori-Jesu 2008). Antioxidants maintain the structure through the reduction of free-radicals, thereby protecting the product against deleterious effect of highly reactive oxygen (Rimbach et al 2002).

There are two types of antioxidants that can be used to retard or prevent oxidative deterioration namely, synthetic and natural antioxidants (Zhang et al 2010). The use of synthetic antioxidants, such as butylated hydroxyanisole (BHA), butylated hydroxytoluene (BHT) and tert-butyl hydroquinone (TBHQ), has come under scrutiny as they are said to contain toxic components that result in health risks (Zhang et al 2010). This caused a restriction on the use of synthetic antioxidants in many food industries, thus increasing the demand for more natural antioxidants (Prior 2004). Even though natural antioxidants are more expensive than their synthetic counterparts, they remain a safer option to use in food products (Prior 2004). Natural oxidants (e.g. tocopherols and tocotrienols) can either be made up of phenolic or polyphenolic compounds (e.g. tannic acid and ellagitannin), with their antioxidant activities depending on structure, hydrogen-donating ability and ability to chelate metal ions (Babovic et al 2010).

Low oxidative stability in relation to saturated fatty acids occurs due to the hydrogenation of unsaturated oils (Kamal-Eldin and Anderson 1997). Regular sunflower oil has high levels of polyunsaturated fatty acids, such as linoleic acid, and is mainly composed of triacylglycerols (98-99%) (Shahidi 2005). Lipid oxidation is mainly influenced by the degree of unsaturated fatty acids present within a product (Frega et al 1999). Studies by Yanishlieva and Marinova (1995) showed that the binding of fatty acids and triacylglycerols affect the lipid substrate through oxidation when no antioxidants are present. The work done by Shmoulvich (1994) has suggested a correlation between tocopherols and fatty acids in vegetable oils. However, no full correlation(s) between the different tocopherol derivatives and fatty acid composition in sunflower oil have been done yet.

Aim

The aims of this chapter are to test various oil extraction methods to obtain measurable tocopherols and amino acid profiles from sunflower seeds. These will then be used to determine whether there is any correlation between *α, β, γ* and *δ* tocopherol levels and the fatty acid (oleic/linoleic) composition of different sunflower accessions analysed.

Objectives

The aims of the chapter will be addressed using three different objectives:

1. the testing of current protocols, and/or the developing of a suitable protocol, to extract *α, β, γ* and *δ* tocopherol from sunflower seeds which are compatible with GC-MS/MS sample analyses;

2. the evaluation of different GC-MS/MS protocols and the required specific spiking levels for the internal calibration curves for *α, β, γ* and *δ* tocopherols; and

3. the determination of the oleic and linoleic acid profiles for the sunflower seed accessions under evaluation.

3.3 Materials and methods

3.3.1 Seed material

One hundrered and four sunflower accessions were tested in this study, with accessions listed in table 3.1A (Appendix 3). A total of about 50 g seeds for each of the 104 accession were obtained from the Agricultural Research Council's – Grain Crop Institute (ARC-GCI) to extract tocopherols from each sample. These accessions were planted at ARC-GCI's experimental farm, Potchefstroom, during the summer months of November and December 2013. The seeds were harvested and collected in March 2014.

3.3.2 Seed fatty acid analysis

3.3.2.1. High-resolution nuclear magnetic resonance (NMR) analysis

A total of 12 g of seeds were used to analyse the seed oil content using a SLK SG100 benchtop spectrometer (Spinlock, Magnetic Resonance Solutions), with software version 3.0. This benchtop machine automatically and simultaneously determined the seed moisture, total oil, and oleic acid, all without destroying the seeds. This non-destructive NMR technology allowed three independent measurements of each sample. This confirmed the oleic acid and total oil content of each accession. Three independent measurements of each seed batch were averaged and used in further statistical analyses.

3.3.2.2 Near infrared analysis

Seeds to be analysed were poured into an open Perten dish (170 mL) and excess seeds were removed as recommended by the manufacturer. The dish with its seed content was then analysed in a Perten DA 7250 Near Infrared Analyzer using the Perten Sunflower Oil Analyses algorithm. In short, the seeds are scanned 240 times using the NIR technology while the dish is

slowly rotated. This measures the top layers of seeds across the dish and uses the correction of the algorithm analysed and provides the average from the different components measured during the readings. This procedure was repeated three times for each seed batch and the average were used for statistical analyses. Similarity to the spinlock, the DA 7250 was used to determine the oleic acid, linoleic acid, moisture content and oil without destroying the seeds, with the distinction that this instrument uses near infrared detection technology and not NMR.

3.3.3 Preparation of solutions for tocopherol standards

All four tocopherol derivatives were prepared in hexane. Stock solutions (1000 ppm) of *α*, *β, γ,* and *δ* tocopherol (Merck) were prepared in 100 mL volumetric flasks by dissolving 10 mg of each in, hexane. A 100 ppm subsequent dilution of *γ* and *δ* were made from these, using hexane.

3.3.4 Tocopherol extraction

Seeds obtained from ARC-GCI were moved to the laboratory where they were crushed using an IKA A11 basic laboratory grinder until it is a fine powder with uniform consistency formed. The powered seed material (0.5 g) was used with various extraction solvents, including toluene, ethanol, 50% dichloromethane/hexane and isopropanol, to determine the best option for GC MS/MS analyses of all four tocopherols. Pure hexane worked best for the GC MS/MS analysis for all four tocopherol derivatives. The powered seeds (0.5 g) were also used for tocopherol extraction and spiked at various concentrations (see section 3.4.2: vitamin E standard calibration curve GC MS/MS). Each sample was supplemented with 2.5 mL ethanol, 0.25 mL potassium hydroxide (600 g/L) and 100uL ascorbic acid (to prevent oxidation of tocopherols). The spiking of each sample was done on the dry powdered seeds. Samples were vortexed (PRO VSM-3 Vortex mixer) for 10 s at 2500 rpm. All votexing was conducted under these conditions unless otherwise indicated. After incubation (15 min in water bath at 80° C), vortexing steps were repeated twice, before samples were cooled for 10 min on ice, with the immediate addition of 2.5 mL

sodium chloride (10 g/L). The samples were again vortexed (10 s) and tocopherols were extracted with the addition of 2.5 mL n-hexane-ethyl acetate (9:1 v/v) and vortexed for 15 s. The entire volume was centrifuged (Beckman Coulter, Allegra X-22R centrifuge) for 5 minutes at 5800 rpm after which the organic layer was removed and placed in a new 15 mL falcon tube. The seed residues were re-extracted four times with 2.5 mL n-hexane-ethyl acetate (9:1 v/v) as described above. The total extracted tocopherol samples (11 - 12.5 mL) were placed in a nitrogen gas dryer (TurboVap LV, Biotage) until dried (40-60 minutes) to combine the extracted layers. A white pellet was retrieved and reconstituted with 2 mL absolute ethanol. The 2 mL tocopherol and ethanol solution was filtered through a 3 cc/mL Terumo syringe filter.

3.3.5 Gas Chromatography and Tandem Mass Spectrometry (GC MS/MS) analyses

Ethanol saponified samples were analysed on a Hewlett Packard 6890 GC system with Micromass Quattro micro MS/MS (WATERS, USA) using Masslynx software (version 4.1). A Zebron ZB-MultiResidue 2 capillary GC column (0.20 µm × 0.25 mm × 30 m; Phenomenex, USA) was used for separation. Split injections of 2 µL were made using the carrier gas (UHP helium) with a flow rate set at 1 mL/min. The GC interface temperature was 250°C and the source temperature was 180°C. The initial temperature was set to 70°C, held for 2 min, followed by ramps to 220°C at 20°C.min^{-1}, 260°C at 5°C.min^{-1} and 300°C at 20°C.min^{-1}. This was concluded by a hold for 15 min at 300°C. The final isothermal hold time was 28 min at 300°C.

The mass spectrometer was operated in MRM-scan (multiple reaction monitoring) mode with the transitions for the respective tocopherols listed in the table 3.1. Quantification was performed using the summed total ion counts (TIC) of all the MRM signal intensities.

Table 3.1: GC MS/MS transition times for the different tocopherols: Tocopherols were detected with the MRM-scan mode (Multiple Reaction Monitoring), thus allowing the selective detection of target compounds even in the presence of large amounts of co-extracted interferences.

Compound	MRM1 (Collision energy) (V)	MRM2 (V)	MRM3 (V)	Retention time (Minutes)
***α*-Tocopherol**	430.0 – 57.0 (30)	430.0 – 165.0 (20)	430.0 – 205.0 (25)	23.20
***β*-Tocopherol**	416.0 – 151.0 (10)	416.0 – 191.0 (20)	-	22.13
***γ*-Tocopherol**	416.0 – 151.0 (10)	416.0 – 191.0 (20)	-	22.32
***δ*-Tocopherol**	402.0 – 137.0 (20)	402.0 – 177.0 (10)	-	21.30

3.3.6 Statistical analyses

The results obtained was evaluated by analysis of variance (ANOVA) and means separated by RStudio (version 0.99.484, © 2009-2015 RStudio, Inc. USA) (http://www.rstudio.org/) to determine the correlation significances using the ggplot2, reshaspe2 and lattice respectively. RStudio was used due to its ability to run several correlation tests at once using large amounts of data on an open source integrated development environment.

3.4 Results

3.4.1 Seed fatty acid analysis (Spinlock NMR and DA 7250 NIR analyzer)

The moisture content was measured for all 104 sunflower samples and two accessions (KP 417 and KP 1226) had high moisture content of 14% (which was due to the seeds not being properly dried) followed by AGSUN 5279 at 7% (figure 3.1). The rest of the 101 samples had moisture percentages of 4% and lower. More than 90% of the 104 sunflower accessions had higher levels of linoleic acid compared to oleic acid (figure 3.3). This was expected since sunflower contains of high levels of polyunsaturated fatty acids. The accession KP 417 consists of the highest level of both linoleic and oleic acid. Low concentrations of linoleic acid were found in 8 commercial lines (18,12 – 50%) with the lowest observed for SY 4045 (18,12%). The rest of the 12 commercial lines and 84 sunflower accessions had mid to high concentrations (>50%) of linoleic acid. All commercial lines had low concentrations of oleic acid and ranged between 13% and 25%. The lowest concentration of oleic acid was observed in the commercial line PAN 7100 which contained 13%. There were only two sunflower accessions that had low oleic acid concentrations under 25% namely, KP 1608 (17%) and KP 1069 (23%). Out of the 84 sunflower accessions only 28 accessions had mid to high oleic acid (50% and above). The rest of the sunflower accessions (56) had low concentrations and ranged between 13% and 49%.

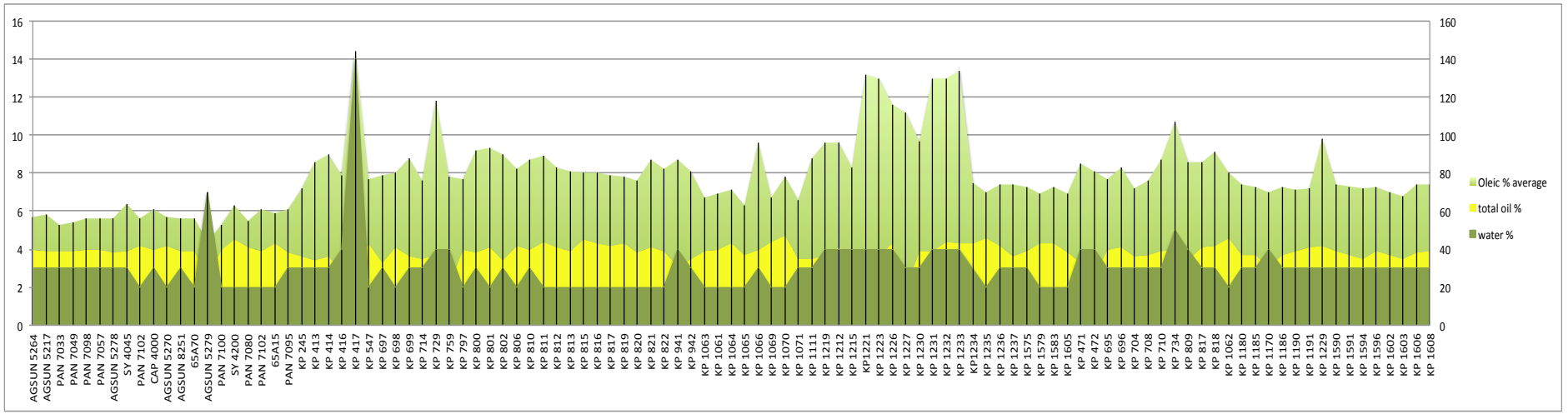


Figure 3.1 Measurement of total oil, water and average oleic acid: Sunflower seed analyses for total oil (yellow), water (dark green) and oleic percentages (light green) as measured by Spinlock. Water percentages for most samples ranged between 2-7% and is inline with percentages expected for samples in storage. KP 417 and KP 1226 had the highest moisture percentage at 14%. The total oil percentage for all the samples ranged between 19 and 47%. This is expected since many of these lines are used in breeding mid to higher oil content cultivars, or cultivars themselves. A few samples had extremely high oleic acid percentages above the 100 % calibration curve. This may be due to high moisture content (e.g. KP417) that is skewing measurements, or simply because the lines with high oleic acid content are outside of the current standard calibration curve.

3.4.2 Vitamin E standard calibration curve (GC MS/MS)

Samples were spiked with all four tocopherol derivatives at different concentrations to obtain an internal standard calibration curve (tables 3.2 and 3.3). The USDA National Nutrient Database for Standard Reference (SR) (http://fnic.nal.usda.gov/food-composition/vitamins-and-minerals) determines the levels of each tocopherol present across various seed crops. According to the USDA for standard reference *α* and *β* tocopherol in dried seeds had a higher concentration than *γ* and *δ* tocopherol. Therefore higher spiking concentrations for *α* and *β* tocopherol were used than for *γ* and *δ* tocopherol. Additionally, it was observed that lower concentrations for *α* and *β* tocopherol caused the internal standard antioxidants to be unstable.

Table 3.2 Internal standard concentrations for *α* and *β* tocopherols: The blank samples (blank 1-3 consist of bird sunflower seeds) represent analyte free matrixes that are used to evaluate if any contamination occurred during the complete extraction procedure (http://www.esslaboratory.com/ pdf/du.pdf). The spiking levels (level 1-7 consist of bird sunflower seeds) are interference free matrixes spiked with known concentrations of tocopherol derivatives. This is done to determine without sample matrix, if the procedure is working within the established control limits (http://www.esslaboratory.com/pdf/du.pdf). The recovery of the spiked levels, i.e. the average of the blank minus each spiking level, is used to determine the accuracy in the given matrix. The 104 sunflower accessions were tested 3 times for statistical analyses except the blank and spiked samples. A series of concentrations of 0, 25, 50 100, 150, 200, 300 and 400 ppm of *α* and *β* tocopherol were prepared for the development of a standard curve.

Samples	Final concentration (ppm)	Spiking volume (µL)	Standard (ppm)
Blank 1	0	0	
Blank 2	0	0	
Blank 3	0	0	
Level 1	25	12.5	1000
Level 2	50	25	1000
Level 3	100	50	1000
Level 4	150	75	1000
Level 5	200	100	1000
Level 6	300	150	1000
Level 7	400	200	1000
Control 1	50	25	1000
Control 2	300	150	1000

Table 3.3 Internal standard concentrations of *γ* and *δ* tocopherols: A series of concentrations for *γ* and *δ* tocopherol were prepared for 0.5 g-grounded sunflower seed material to obtain a linear internal standard calibration curve.

Samples	Final concentration (ppm)	Spiking volume (µL)	Standard (ppm)
Blank 1	0	0	
Blank 2	0	0	
Blank 3	0	0	
Level 1	0.5	5	100
Level 2	1	10	100
Level 3	1.5	15	100
Level 4	2	20	100
Level 5	4	40	100
Level 6	8	80	100
Level 7	10	100	100
Control 1	1	10	100
Control 2	8	80	100

3.4.3 Fatty acid and tocopherol analyses

The oleic acid levels were measured for the samples where the high, mid and low oleic acid levels were determined. A total of 8 samples had high oleic (80% and above) contents and 11 samples contained mid oleic (55% - 75%) acid (figure 3.2) (http://www.sunflowernsa.com/oil/what-is-high-oleic-sunflower-oil/). The first twenty samples represent commercial lines. All of these samples had low oleic acid levels (below 25%) but a higher percentage of total oil. This correlates with farmers in South Africa being paid on the total oil content rather than oil type.

Between the two main fatty acids analysed there were higher levels of linoleic acid in all the 104 sunflower accessions compared to the oleic acid levels (figure 3.3). Within the sunflower accessions a major difference was distinguished between the lowest concentrations of oleic acid (13% - Pan 7100) and the highest for linoleic acid (283,87% - KP 417) (figure 3.3). The sample KP 417 also contained the highest percentage of oleic acid (113%) and moisture content (14%) but in reverse contains one of the lowest concentrations of *α* tocopherol (192,8 ppm).

The results indicate that there's a relation between oleic and linoleic acid among a few lines e.g. KP 714 had low oleic (36.46%) acid and high linoleic (81%) acid but a reverse relationship was observed for KP 729 which consisted of 81% oleic acid and 36,46% linoleic acid, similar observations were made for the following samples namely, KP 734, KP 800, KP 1111, KP 1219, KP 1221, KP 1223, KP 1226, KP 1227, KP 1231, KP 1232 and KP 1233 respectively (figure 3.3).

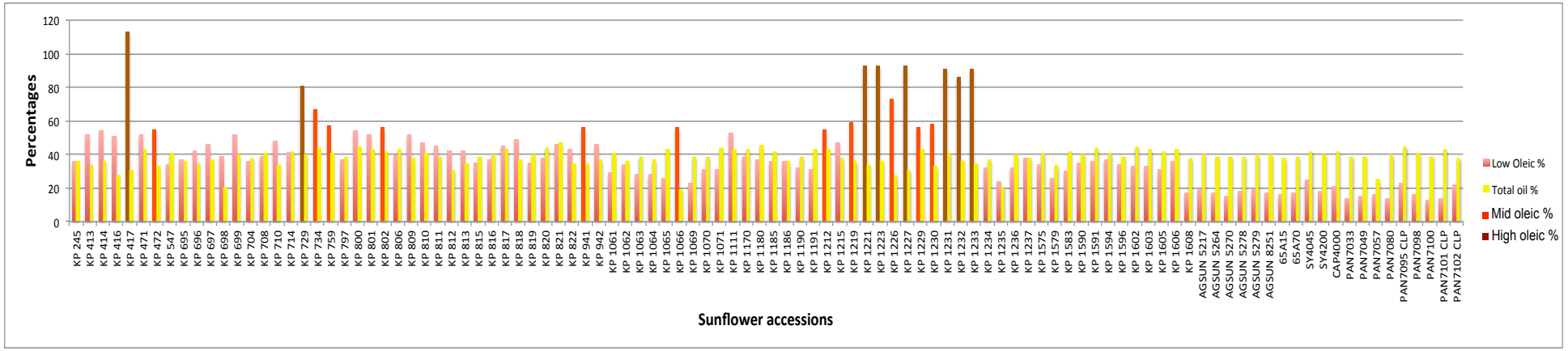


Figure 3.2: Oleic acid and total oil measurements: The high-resolution nuclear magnetic resonance (NMR) was used to determine the seed total oil (yellow), and oleic acid levels without destroying the seeds. Each sample was measured three times and the average was used as percentages for both oleic acid (different shades of red) and total oil. The different oleic acids levels ranges from <55 - >75 (high – maroon [>75], mid - red [55-75] and low - orange [<55] oleic acid levels). The total oil percentage for each accession consist of different fatty acids i.e KP 472 have 33% total oil but the oil is made up of 55% oleic and 57,3% linoleic acid.

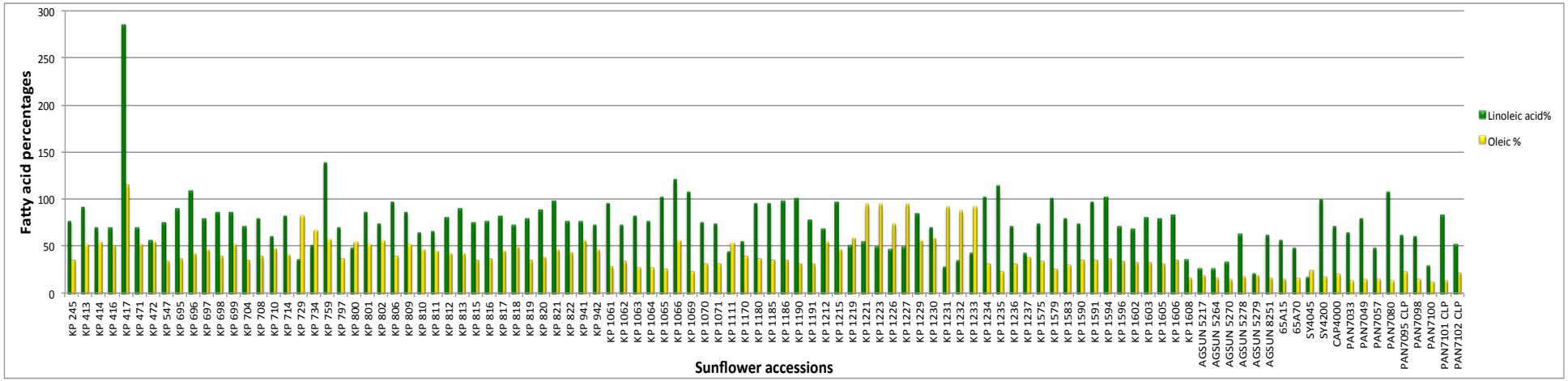
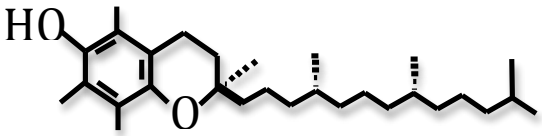
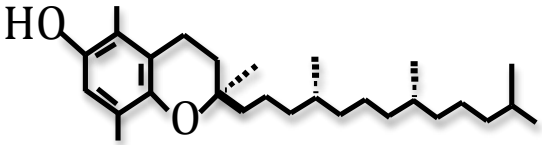
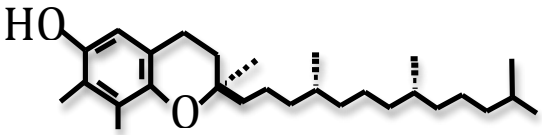
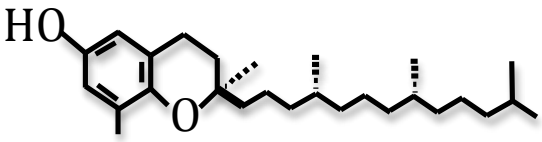


Figure 3.3: Comparison between oleic and linoleic acid: The Near infrared (DA 7250 NIR Perten) was used to measure linoleic acid (green) and the high-resolution nuclear magnetic resonance (Spinlock) for the oleic acid (yellow). Both of these procedures were repeated three times for each seed batch and the average were used as percentages. The results indicate that higher levels of linoleic acid were found in most of the samples compared to oleic acid. A few of the samples tested on the DA 7250 have percentages higher than the calibrated value (100%) this is due to the high water percentages (14%) for these specific lines, as it interferes with the fatty acid composition and lead to false readings.

The composition of all four tocopherol derivatives was profiled using the GC MS\MS (table 3.4). *α*-Tocopherol was measured as one of the highest among all four tocopherol derivatives in all 104 sunflower accessions (figure 3.4). A wide variation was observed between all four tocopherol derivatives (figure 3.4). *α*-Tocopherol concentrations ranged between 171.15 and 697.71 ppm for all sunflower accessions, while *β, γ* and *δ* tocopherol levels ranged between 33 – 104.58 ppm, 0.38 – 5.1 ppm and 0.31 – 2.68 ppm respectively, for per 0.5 g seed material (table 3.1A, Appendix 3). Among all four tocopherol derivatives *α*-tocopherol was the predominant tocopherol and contained a mean average value of 331.41 ppm across all 104 sunflower accessions (table 3.1A, Appendix 3). The highest levels of total tocopherol were observed in the sample KP 1237 with a mean value of 742.99 ppm (table 3.1A, Appendix 3). The sunflower accession with the highest *α* tocopherol was KP 1237, which had a concentration of 697.71 ppm, whereas KP 810 (104.58 ppm) had the highest concentration for *β* tocopherol, KP 417 (5.13 ppm) had the highest *γ* tocopherol concentration and KP 714 (2.68 ppm) had the highest *δ* tocopherol concentration (table 3.1A, Appendix 3)

The accession KP 1237 might contain the highest total and *α* tocopherol but had mid concentrations of linoleic acid (74%), low levels of oleic acid (34%) and high levels of total oil (41%). The same goes for KP 810, which have high levels of *β* tocopherol but low levels of oleic acid (47%), mid levels of linoleic acid (65.75%) and high levels of total oil (41%). This indicates that these specific lines were bred for total oil and not oil content. The sample KP 714 had the highest levels of *δ* tocopherol as well as high levels of linoleic acid (82,98%), and total oil (42%) but low levels of oleic acid (41%).

Table 3.4 The complete profiles of *α, β, γ* and *δ* tocopherol confirmed by the GC MS/MS: *α, β, γ* and *δ*-Tocopherol are structurally similar but consist of changes in the chromanol head. The profile scan of *β* and *γ* tocopherol indicate that these two derivatives consist the exact same mass and formula. The retention time indicates that *β* and *γ* tocopherol are separated by 19 seconds, making them very close to one another but still able to differentiate between the two derivatives. *α* and *δ*-Tocopherol indicated good peak shape and separation with a retention time of 23.20 min for *α* tocopherol, and 21.30 min for *δ* tocopherol.

Form	Structure	Mass (g)	Formula	Retention time (Min)
***α*-Tocopherol**		430.381073	$C_{29}H_{50}O_2$	23.20; 430.00
***β*-Tocopherol**		416.365417	$C_{28}H_{48}0_2$	22.13; 416.00
***γ*-Tocopherol**		416.365417	$C_{28}H_{48}0_2$	22.32; 416.00
***δ*-Tocopherol**		402.3497	$C_{27}H_{46}O_2$	21.30; 402.00

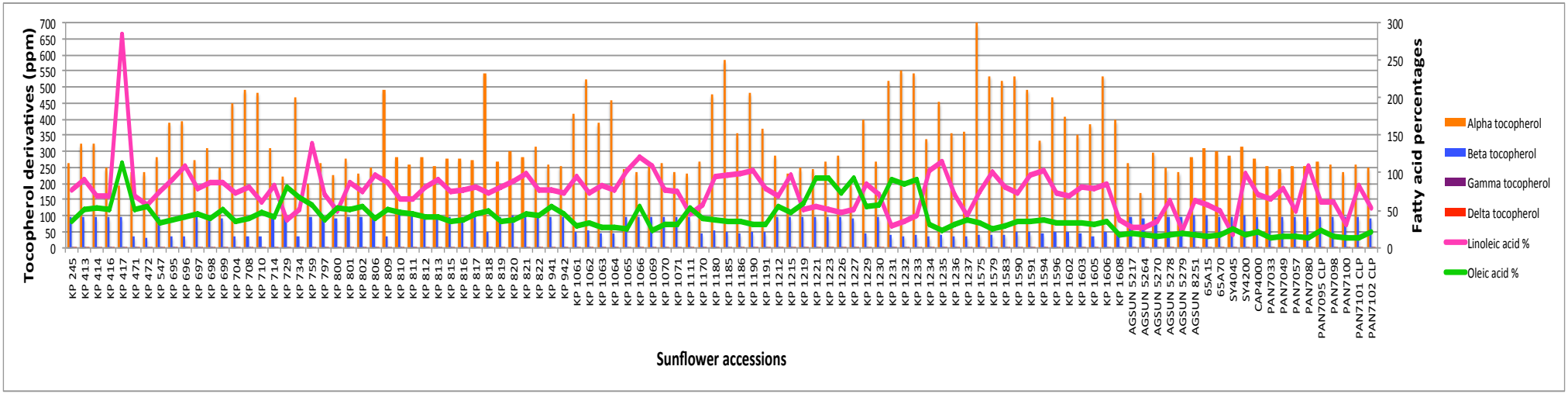


Figure 3.4 Different levels of *α, β, γ* and *δ* tocopherol in comparison with the fatty acid profiles: All four tocopherol derivatives were determined by GC MS/MS and were detected with the MRM-scan mode (Multiple Reaction Monitoring). Each of the 104 samples for each tocopherol derivative was repeated three times for statistical analysis. This graph indicates the different levels in parts per million (ppm) for *α* (orange), *β* (yellow), *γ* (purple) and *δ* (red) tocopherol as well as oleic (green) and linoleic acid (pink) present in each of the 104 samples. The tocopherol derivatives ranged between 0.3 – 697 ppm and were plot against the oleic and linoleic acid (18,12% – 283,87%). The accessions KP 417 had the highest percentage of linoleic (283,87%) and oleic acid (113%). This is due to the 14% water content found in this sample, as the water percentage influence the fatty acid composition.

A correlation test between all four tocopherol derivatives, total tocopherol, oleic and linoleic acid indicated that *α* tocopherol was highly correlated with total tocopherol (r=0.89). This is expected since *α* tocopherol is the major contributor of the tocopherols to the total tocopherol in sunflower accessions analysed (table 3.5). A positive correlation was established between *α* and *β* (r=0.93), *α* and *γ* (r=0.88), *β* and *γ* (r=0.85) and *γ* and *δ* (r=0.72) respectively (table 3.5). The high significant positive correlation between *α* and *β* tocopherol suggests that an increase in *α* tocopherol will result with an increase in *β* tocopherol. Total tocopherol was significantly positively correlated with *α* (r=0.89), *β* (0.84), *γ* (r=0.76) and *δ* (r=0.5) (table 3.5). No correlation was found between linoleic acid and *α, β, γ* and *δ* tocopherol. The results indicate that there is a positive correlation between oleic acid, and *γ* (r=0.17) and *δ* (r=0.23) tocopherol but none between oleic acid, *α* and *β* tocopherol (table 3.5).

Table 3.5: The correlations between *α, β, γ, δ* and total tocopherol, as well as oleic and linoleic acid: This was calculated in RStudio using the ggplot2, reshaspe2 and lattice packages to determine the correlation significances among the fatty acids and the four tocopherol derivatives.

	***β* tocopherol**	***γ* tocopherol**	***δ* tocopherol**	**Total tocopherol**	**Oleic acid**	**Linoleic acid**
***α* tocopherol**	0.93**	0.88**	0.62*	0.89**	0.043^{NS}	0.0084^{NS}
***β* tocopherol**		0.85**	0.63*	0.84**	-0.0011^{NS}	0.043^{NS}
***γ* tocopherol**			0.72**	0.76**	0.17 *	0.071^{NS}
***δ* tocopherol**				0.5*	0.23*	-0.024^{NS}
Total tocopherol					-0.0051^{NS}	-0.0046^{NS}

* Low to moderate linear relation,

** High linear relation,

^{NS} No significance

3.5 Discussions

Sunflower oil contains mainly four fatty acids, namely stearic, palmitic, oleic and linoleic acid (Lacomble and Berville 2001). Of these, sunflower oil consists mainly of linoleic and oleic which make up 90% of the total fatty acid content (Škorić et al 2008), therefore only oleic acid, linoleic acid and total oil was considered for analyses and comparisons here. Oil content is very important as it determines the commercial value of a crop on the international markets (Shahidi F 2005), however in South Africa, total seed oil percentage is still largely used for seed price determinations. The total seed oil was measured for all 104 samples and ranged between 19% and 47% (figure 3.2). There were six accessions with total seed oil below 30% namely, KP 1066 (19%) followed by KP 1235 (21%), KP 698 (21%), PAN 7057 (25%), KP 1226 (28%) and KP 416 (28%) (figure 3.2). The rest of the 98 samples had total seed oil higher than 30% with the highest percentage found in KP 821 (47%). In a comparison to the fatty acid content this accession (KP 821) had high levels of linoleic acid [>75] and low levels of oleic acid [<55]. The accession with the lowest total seed oil, KP 1066, consists of high levels of linoleic acid [>75] and mid levels [55 - 75] of oleic acid.

Once oil seeds have been harvested, factors such as moisture affect the value of a crop when it comes to storage (Harrington 1959). Harrington (1959) found that a high seed moisture percentage lowered the quality of the seed since insects and fungal growth was supported, thus making it unsafe for future use. Studies shows that seed moisture content between and below 4-8% are safe and have no threat of insect or fungal growth (Harrington 1959). Hofman and Hellewang (1997) indicated that the maximum moisture content for sunflower seeds is 8% for short time storage (1-3 years) and 9% for long time storage (4-6 years). This indicates that seeds with high moisture content may influence the oil quality due to fungal and insect contamination. Harrington (1970) suggested that oily seeds would usually tolerate drying to a lower moisture content compared to non-oily seeds. Thus seeds with moisture content between and below 6-8% can be used for long-term storage for

genetic conservation indicating that moisture content is an important factor in determining seed longevity. In this study we found two samples with moisture content higher then 10% (e.g. KP 417 and KP 1226). The sample KP 417 had high levels of oleic (113%) and linoleic acid (283,87%) due to the high percentage of moisture (14%). The sample KP 1226 also had high percentage of moisture (14%) as well as high levels of oleic acid (93%). These two samples are the only samples with high moisture levels and should be dried and repeated in future with a maximum of 10% moisture content. The results found in this study emphasise the importance of moisture in oil seeds and how it affects the fatty acid content because according to the manufacturer's of the Perten and Spinlock, samples with a moisture content higher then 10% will manipulate the results.

A comparison between oleic and linoleic acid indicated higher levels of linoleic acid for most lines. This suggests that sunflower has higher levels of polyunsaturated fatty acid making it more susceptible to oxidation. From the 104 lines the results indicate that all the 20 commercial lines had low oleic acid, but high total oil content, thus suggesting that these commercial lines were bred for total oil yield and not oleic acid concentrations (figure 3.1). This would be in line with South Africa farmers receiving payment on sunflower oil percentage (total oil) compared to oil composition. Therefore lines with high yields, are selected rather than high oleic acid lines.

Various studies found a correlation between tocopherol and the fatty acid composition in vegetable oil (e.g rapeseed, sybean, sunflower etc.). Kamal-Eldin and Anderson (1997) found a positive correlation between linoleic acid and α tocopherol ($r = 0.549$), and suggested a positive correlation between γ tocopherol and linolenic acid. A study done by Onemli (2012) found a negative correlation between oleic acid and linoleic acid ($r = -0.99$) but motivated that the correlation was due to climate change. Dolde et al (1999) found that altering the fatty acid content in sunflower did not affect the total tocopherol concentrations. They've also indicated that there is no correlation between the tocopherol concentration and the degree of unsaturated fats in sunflower seeds with genetically altered fatty acid composition. Here we've

found that *α* tocopherol is the predominant tocopherol and demonstrated a positive correlation with total tocopherols. A positive correlation was also observed between tocopherol derivatives (*α, β, γ* and *δ*) and total tocopherol. There was no significant correlation found between the fatty acids and tocopherol derivatives, it only showed a very low correlation between *γ* and oleic acid, and *δ* and oleic acid. These results are therefore in good agreement with the report of Dolde et al (1999), suggesting that there is no significant correlation between the two homologues (tocopherol and fatty acid) in sunflower seeds.

The high positive correlation was found between *α, β, γ and δ* tocopherol derivatives and as well as between *α, β, γ, δ* and total tocopherol. The highest correlation was observed between *α* and *β* tocopherol (0.93), followed by a correlation between *α* and *γ* tocopherol (0.88), *α* and total tocopherol (0.98), *β* and *γ* tocopherol (0.85), *β* and total tocopherol (0.84), *γ* and *δ* tocopherol (0.72) and lastly *γ* and total tocopherol (0.76) (figure 3.4). A low to moderate correlation was found between *γ* and oleic acid, and *δ* and oleic acid. This suggests that improving the fatty acid content of sunflower would be much more effective than changing the tocopherol concentrations to obtain oil less prone to oxidation. Considering that almost no correlation was found between all four tocopherol derivatives, oleic and linoleic acid (figure 3.4), breeding efforts should focus on improving the fatty acid content of these genotypes to mid-fatty acids and high tocopherol concentrations to produce sunflower oil less prone to oxidation.

3.6 Conclusions

α-Tocopherol is one of the main tocopherols present in sunflower oil, and normally represents up to 90% of the total tocopherol content (Hu et al 2010). A tocopherol extraction protocol that is GC MS/MS compatible was successfully developed and tested on sunflower seeds. The specific spiking levels for each tocopherol derivative were successfully determined. This was done to create a standard calibration curve and to determine if the procedure is working within the established control limits for each tocopherol derivative.
The results obtained from the GC MS/MS indicate that *α* tocopherol was the predominant form among the four tocopherol derivatives ranging from 171 - 697ppm (table 3.1A, appendix 3) in the samples tested. *β*-Tocopherol ranged between 33 - 104ppm whereas *γ* and *δ* tocopherol were present in low concentration (0.31 - 5.1ppm) (table 3.1A, appendix 4). Warner 2005 show that sunflower oil stability can be improved by decreasing linoleic acid and also by increasing the proportions of *γ* and *δ* tocopherol. Breeders have been focusing on the partial replacements of *α* tocopherol (*β*, *γ*, *δ* and other tocopherols) since *α* tocopherol is found to be the weakest antioxidant *in vitro* (Hu et al 2010).

High levels of linoleic acid were observed in approximately 90% of the sunflower accessions tested when compared to oleic acid. The sample KP 417 and KP 1226 had the highest moisture levels (14%) which resulted in high levels of oleic and linoleic acid but needs to be reanalysed since a moisture content higher then 10% influence the fatty acid profiles and give inaccurate readings. *α*-Tocopherol was highly correlated with total tocopherol (r = 0.89) indicating that the contribution of this specific tocopherol is highly important among all four tocopherol derivatives. A positive correlation was observed among the *α, β, γ and δ* tocopherol derivatives. A high significance was also observed between *α, β, γ, δ* and total tocopherol. No significance was observed among linoleic acid and all four tocopherol derivatives but a positive correlation was observed between oleic acid, *γ* (r =0.17) and *δ* (r =0.23) tocopherol. Therefore according to the results obtained, tocopherol

concentrations do not seem to influence either of the fatty acids, thus breeding can be done for both tocopherol and fatty acid concentrations independently of each other (Hu et al 2010).

References

Aluyor EO and Ori-Jesu M 2008. The use of antioxidants in vegetable oils – A review. *African Journal of Biotechnology* 4837-4841

Ammari F, Cordella CBY, Boughanmi N and Rutledge DN 2012. The increase in oxidative stability of sunflower oil enriched with *Nigella sativa* L. seed extracts. *Journal of Food Measurement and Characterization* 12-20

Azizkhani M, Kamkar A and Mozaffari Nejad AS 2011. Effects of tocopherols on oxidative stability of margarine. *Journal of the Chemical Society of Pakistan* 134-137

Baldwin WS, Minneapolis and Keeney KW 1974. Methylation of tocopherols. *United States Patent Office* 3819657

Bandarra NM, Campos RM, Batista I, Nunes ML and Empis JM 1999. Antioxidant synergy of *α*-tocopherol and phospholipids. *Journal of the American Oil Chemists Society* 905-913

Burton GW and Ingold KU 1981. Autoxidation of biological molecules 1. The antioxidant activity of vitamin E and related chain-breaking phenolic antioxidants *in vitro*. *Journal of the American Oil Chemists Society* 6472-6477

Department of Agriculture Forestry and Fishery 2011. Sunflower seed market value profile, chain (2010-2011). 204-232

Dieffenbacher and Pocklington 1992. Standard methods for the analysis of oils, fats and derivatives. *1St Supplement to the 7th Revised and Enlarged Edition* 1-7

Evans HM and Bishop KS 1922. Fetal resorption. *Science* 650

Fernandez-Martinez JM, Perez-Vich B and Velasco L 2009. Mutation breeding for oil quality improvement in sunflower. Induced plant mutations in the genomics era. Plant breeding section, joint FAO/IAEA division of nuclear techniques in food and agriculture, *International Atomic Energy Agency* 177-181

Hamilton RJ, Kula C, McNeill GP, Padley FB and Pierce JH 1998. Effects of tocopherols, ascorbyl palmitate, and lecithin on autoxidation of fish oil. *Journal of American Oil Chemists' Society* 813-823

Hass CG, Tang S, Leonard S, Traber M, Miller JF and Knapp SJ 2006. Three non-allelic epistatically interacting methyltransferase mutations produce novel tocopherol (vitamin E) profiles in sunflower. *Theoretical and Applied Genetics* 767-782

Herrera E and Barbas C 2001. Vitamin E: action, metabolism and perspectives. *Journal of Physiology and Biochemistry* 43-56

Hofius D and Sonnewald U 2003. Vitamin E biosynthesis: biochemistry meets cell biology. *Trends in Plant Science* 6-8

Hu J, Seiler G and Kole C 2010. Genetic, genomics and breeding of sunflower. *CRC press, Enfield, USA* 1-40

Hunter SC and Cahoon EB 2007. Enhancing vitamin E in oilseeds: Unraveling tocopherol and tocotrienol biosynthesis. *Lipids* 97-108

Ikeda N and Fukuzumi K 1977. Synergistic antioxidant effect of nucleic acids and tocopherols. *Journal of American Oil Chemists' Society* 360-366

Jorge N, Márquez-Ruiz G, Martín-Polvillo M, Ruiz- Méndez MV and Dobarganes MC 1996a. Influence of dimethylpolysiloxane addition to edible oils: dependence on the main variables of the frying process. *International Journal of Fats and Oils* 14-19

Jung MY and Min DB 1990. Effects of *α, γ* & *δ* tocopherols on oxidative stability of soybean oil. *Journal of Food Science* 1464-1465

Kilcast D and Subramaniam P 2000. The stability and shelf life of food. *Woodhead Publishing Limited and CRC press LLC, Cambridge, England* 1-19

Lacombe S and Berville A 2001. A dominant mutation for high oleic acid content in sunflower (*Helianthus annuus* L.) seed oil is genetically linked to a single oleate-desaturase RFLP locus. *Molecular Breeding* 129-137

Lea and Ward 1959. Relative anti-oxidant activities of seven tocopherols. *Journal of The Science of Food and Agriculture* 537-548

Matrix spike - ESS Laboratory (http://www.esslaboratory.com/pdf/du.pdf)

National Sunflower Association (http://www.sunflowernsa.com/oil/what-is-high-oleic-sunflower-oil/)

NDA 2015. "Abstract of agricultural statistics", National department of agriculture, Pretoria, South Africa, 2015

PANNAR Seeds 2006. "Sunflower production guide", Pannar Seeds Production Guide Series

Prior RL 2004. Absorption and metabolism of anthocyanins: potential health effects. In Meskin M, Bidlack WR, Davies AJ, Lewis DS and Randolph RK *Phytochemicals: mechanisms of action. Boca Raton, FL: CRC Press* 1-19

Rimbach G, Minihane AM, Majewicz J, Fischer A, Pallauf J, Virgli F and Weinberg PD 2002. Regulation of cell signalling by vitamin E. *Proceedings of the Nutrition Society* 415-425

RStudio (2009-2015). RStudio: Integrated development environment for R (Version 0.99.484) [Computer software]. Boston, MA. (http://www.rstudio.org/)

Shahidi F 2005. Bailey's industrial oil and fat products. In: Sunflower Oil. Shahidi F and Grompone MA 665-725

Shmoulvich VG 1994. Interrelation of contents of unsaturated fatty acids and vitamin E in food product lipids. *Applied Biochemistry and Microbiology* 547-551

Škorić D, Jocic S, Sakac Z and Lecic N 2008. Genetic possibilities for altering sunflower oil quality to obtain novel oils. *Canadian Journal of Physiology and Pharmacology* 215-220

Tavassalkar S, Mishra H and Madhavan S 2012. Evaluation of antioxidant efficacy of natural plant extracts against synthetic antioxidants in sunflower oil. *Journal of Food Processing & Technology* 1-5

USDA National Nutrient Database for Standard Reference (SR) (http://fnic.nal.usda.gov/food-composition/vitamins-and-minerals)

USDA Foreign Agricultural Services 2012. Production, Supply and Distribution database. (http://www.fas.usda.gov/psdonline/psdHome.aspx)

van der Merwe R, Labuschagne MT, Herselman L and Hugo A 2013. Stability of seed oil quality traits in high and mid-oleic acid sunflower hybrids. *Euphytica* 1-12

Zhang Y, Yang L, Zu Y, Chen X, Wang F and Liu F 2010. Oxidative stability of sunflower oil by carnosic acid compared with synthetic antioxidants during accelerated storage. *Food Chemistry* 1-15

Appendix 3 (chapter 3)

Table 3.1A: The profiles of oleic and linoleic acid and averages for individual and total tocopherol content in 104 sunflower accessions.

Line ID	Name	Total oil %	Water %	Average oleic acid %	Average linoleic acid %	α-Tocopherol (ppm)	β-Tocopherol (ppm)	γ-Tocopherol (ppm)	δ-Tocopherol (ppm)	Total tocopherol (ppm)
KP245	ND – BLOS	36	3	36	76,9	265,5	96,2	3,8	2,3	367,8
KP413	PSK OVT	34	3	52	91,9	321,8	95,1	3,6	2,3	422,8
KP414	PSK VT	36	3	54	69,9	325,7	96,3	3,8	2,3	428,1
KP416	RK 74 Birdseed	28	4	51	69,9	248,7	94,3	3,2	2,3	348,5
KP417[1]	RK 74-128XRK74-198	31	14[1]	113	283,9	192,9	94,7	5,1	2,4	295,2
KP471	H55-9-2-1-1A	43	2	52	69,8	233,9	33,6	1,1	0,5	269
KP472	H55-9-2-	33	3	55	57,3	233,6	33	0,8	0,3	267,8

Line ID	Name	Total oil %	Water %	Average oleic acid %	Average linoleic acid %	α-Tocopherol (ppm)	β-Tocopherol (ppm)	γ-Tocopherol (ppm)	δ-Tocopherol (ppm)	Total tocopherol (ppm)
	1-1B									
KP547	RK 74-128-1	41	2	34	75,9	280,5	94,8	4,2	2,3	381,9
KP695	H 1185-29	36	3	37	90,4	390,3	36,1	1,8	0,6	428,8
KP696	H 1185-29-1-3-1A	35	3	42	109,3	393,1	35,1	1,4	0,4	429,9
KP697	H 1185-29-1-3-1B	37	4	46	79,9	273,6	94,9	3,2	2,2	373,8
KP698	H 1185-29-1-3-2A	21	4	39	86,3	309,8	94,7	3,2	2,3	410
KP699	H 1185-29-1-3-	40	2	52	86,8	251,4	93,8	3,2	2,3	350,6

Line ID	Name	Total oil %	Water %	Average oleic acid %	Average linoleic acid %	α-Tocopherol (ppm)	β-Tocopherol (ppm)	γ-Tocopherol (ppm)	δ-Tocopherol (ppm)	Total tocopherol (ppm)
KP704	2B H 1185-76-1-2-2B	38	3	36	72,3	450,9	36,2	1,4	0,3	488,8
KP705	H 1185-76-1-2-3	41	2	39	80,3	491,1	36,8	1	0,4	529,3
KP708	H 1185-113-2-3	34	3	48	61,6	483,4	35,1	1,7	0,4	520,6
KP710	H 1185-118	42	2	41	82,9	311,6	96,2	4,1	2,7	414,5
KP714	H1185-142	40	3	81	36,5	221,4	96	2,6	2,6	322,7
KP729	H 1185-200	44	2	67	51,2	466	36,9	2,7	0,4	506,1
KP734	H 1185-248	41	2	57	138,8	199	94,6	2,4	2,3	298,3

Line ID	Name	Total oil %	Water %	Average oleic acid %	Average linoleic acid %	α-Tocopherol (ppm)	β-Tocopherol (ppm)	γ-Tocopherol (ppm)	δ-Tocopherol (ppm)	Total tocopherol (ppm)
KP759	RK 74-24	39	2	37	70,5	264,9	95,1	2,9	2,6	365,5
KP797	PSK 29-31	45	2	54	48,9	228,1	93,2	2,5	2,3	326,1
KP800	PSK 29-37	43	2	52	86,9	275,2	95,4	3,1	2,4	376,1
KP801	PSK 29-45-1A	42	2	56	74,6	231,2	94,3	2,5	2,3	330,3
KP802	PSK 29-45-1B	43	2	40	97,9	250,2	100,1	2,4	2,4	355,1
KP806	SC 3-56A	38	2	52	86,1	491,7	37,2	1,9	0,4	531,1
KP809	SC 3 – 64	41	2	47	65,8	280,2	104,6	3,2	2,5	390,4
KP810	SC 4 – 11A	39	2	45	65,9	257,4	99	2,8	2,4	361,6
KP811	SC 4 – 11B	31	4	42	80,9	282,9	102,4	2,9	2,4	390,7

Line ID	Name	Total oil %	Water %	Average oleic acid %	Average linoleic acid %	α-Tocopherol (ppm)	β-Tocopherol (ppm)	γ-Tocopherol (ppm)	δ-Tocopherol (ppm)	Total tocopherol (ppm)
KP812	SC 4 – 15A	35	3	42	90,8	253	99,2	2,9	2,3	357,4
KP813	SC 4 – 15B	39	2	35	75,6	276,7	96,4	3,3	2,3	378,6
KP815	SC 4 – 24A	40	2	37	77,8	276,8	97,2	3	2,3	379,5
KP816	SC 4 – 24B	43	2	45	82	271,1	95,9	3,1	2,4	372,4
KP817	SC 4 – 42	37	2	49	73,4	543,9	51,9	2,1	0,9	598,9
KP818	SC 4 – 47	40	3	35	80,3	270,2	95,3	3,2	2,3	370,9
KP819	SC 5 – 13	44	2	38	88,9	301,4	99,1	2,6	2,4	405,5
KP820	SC 5 – 43	47	2	46	98,8	282,8	96,3	3,6	2,3	385,1

Line ID	Name	Total oil %	Water %	Average oleic acid %	Average linoleic acid %	α-Tocopherol (ppm)	β-Tocopherol (ppm)	γ-Tocopherol (ppm)	δ-Tocopherol (ppm)	Total tocopherol (ppm)
KP821	SC 5 – 44	35	3	43	77,8	314,4	98,2	3,5	2,3	418,4
KP822	SC 5 – 45	35	3	56	76,9	256,4	97,6	2,9	2,4	359,5
KP941	ND761-10-2-2-1	37	4	46	73,5	255,4	97,8	3,5	2,4	359,1
KP942	ND761-10-2-2-1CMS	41	4	29	95,9	416	50,5	2	1,3	469,8
KP1061	H55-9-2-1-1	36	4	34	73,6	523,5	52,4	1,3	1	578,2
KP1062	H55-9-2-1-1 CMS	39	4	28	83,2	387,8	46,5	1,4	1	436,8
KP1063	H55-9-2-2	37	4	28	77,4	458,2	46,9	1,5	0,9	507,5
KP1064	H55-9-2-	43	4	26	102,9	244,7	96,8	2,6	2,4	346,5

Line ID	Name	Total oil %	Water %	Average oleic acid %	Average linoleic acid %	α-Tocopherol (ppm)	β-Tocopherol (ppm)	γ-Tocopherol (ppm)	δ-Tocopherol (ppm)	Total tocopherol (ppm)
	2 CMS									
KP1065	H55-9-3-1	19	3	56	121,3	236,2	95,8	3,1	2,3	337,5
KP1066	H55-9-3-1 CMS	39	3	23	107,9	251,9	94,8	3	2,5	352,3
KP1069	H52-6-3	39	4	31	76,4	265,1	95,2	2,8	2,4	365,5
KP1070	H52-6-3 CMS	44	4	31	74,4	237,7	94,6	2,5	2,3	337,1
KP1071	H55-2-2	43	4	53	44,9	230,5	94,4	2,6	2,3	329,8
KP1111	RK74-302-2-2	43	3	39	56,2	270,3	46,5	-0,4	0,9	317,4
KP1170	RP869C1-24-1-1	46	2	37	95,5	476,9	52,1	0,5	1,1	530,7
KP1180	RP861C1-87-1	42	3	36	96,1	582,3	50,7	3	1,1	637,1
KP1185	RP941-	36	3	36	98	354,6	45,2	0,7	0,8	401,4

Line ID	Name	Total oil %	Water %	Average oleic acid %	Average linoleic acid %	α-Tocopherol (ppm)	β-Tocopherol (ppm)	γ-Tocopherol (ppm)	δ-Tocopherol (ppm)	Total tocopherol (ppm)
	14-2-1									
KP1186	RP941-18-1-1	39	3	32	101,9	482,4	50,8	1,4	1,1	535,8
KP1190	RP941-29-2-1	43	2	31	78,9	370,2	48,2	0,9	1,1	420,4
KP1191	RP941-29-2-2	43	2	55	69,7	286,8	98,2	3,8	2,3	391,1
KP1212	SRXX3169-1	38	2	47	97	230,6	96,6	2,6	2,3	332,1
KP1215	SRXU3028-2	36	3	59	51,2	248,2	97	2,9	2,5	350,5
KP1219	AP901-56-1-1	34	3	93	55,3	244,6	96,9	3,3	2,4	347,1
KP1221	AP90-56-1-2	36	3	93	50,9	268,7	94,3	3,2	2,4	368,6
KP1223	AP901-	28	4	73	47,8	288,4	95,2	3,6	2,3	389,5

Line ID	Name	Total oil %	Water %	Average oleic acid %	Average linoleic acid %	α-Tocopherol (ppm)	β-Tocopherol (ppm)	γ-Tocopherol (ppm)	δ-Tocopherol (ppm)	Total tocopherol (ppm)
	56-3-4-1									
KP1226[1]	AP901-94-3-2-1 CMS	31	14[1]	93	50,3	197,9	93,1	2,3	2,5	295,7
KP1227	AP901-94-3-2-2	43	2	56	85,7	400,2	46,7	2,1	0,9	450
KP1229	AP901-99-3-3-1	33	3	58	70,8	267,4	94,5	2,9	2,5	367,5
KP1230	AP901-99-3-3-1 CMS	41	2	91	28,6	518,5	38,3	2,5	0,6	560
KP1231	PSK29-45-3	36	3	86	36,1	551,8	37,7	2,1	0,6	592
KP1232	PSK29-45-2	35	3	91	43,9	543,7	38,2	2,9	0,7	585
KP1233	PSK29-	37	4	32	102,3	339,6	34,8	1,8	0,3	376,5

Line ID	Name	Total oil %	Water %	Average oleic acid %	Average linoleic acid %	α-Tocopherol (ppm)	β-Tocopherol (ppm)	γ-Tocopherol (ppm)	δ-Tocopherol (ppm)	Total tocopherol (ppm)
	45-4									
KP1234	H52-6-3	21	4	24	115,4	451,5	40,4	2,7	0,7	495,4
KP1235	H52-6-3 CMS	40	2	32	72	357,3	36,4	2,2	0,4	396,2
KP1236	H55-9-2-1-1 B	38	3	38	43,3	361,5	35,3	1,6	0,4	398,8
KP1237	H55-9-2-1-1 CMS	41	2	34	74,4	697,7	42	2,8	0,5	742,9
KP1579	RP861C1-10-1-1-B	34	3	26	101,1	534,5	41,8	2,2	0,4	578,8
KP1583	RP861C1-87-1-3-B	42	2	30	80	517	41,7	1,5	0,3	560,5
KP1590	RP941-14-2-1-1-B	40	3	35	73,9	531,1	48,4	1,6	0,9	582,1
KP1591	RP941-	44	2	36	97,1	491,5	48,1	1,6	0,9	542,2

Line ID	Name	Total oil %	Water %	Average oleic acid %	Average linoleic acid %	α-Tocopherol (ppm)	β-Tocopherol (ppm)	γ-Tocopherol (ppm)	δ-Tocopherol (ppm)	Total tocopherol (ppm)
	14-2-1-2-B									
KP1594	RP941-18-1-1-2-B	41	2	37	103,2	335,1	45,2	0,6	0,9	381,8
KP1596	RP941-18-1-2-2-B	39	2	34	72,1	465,6	48,4	1,4	0,9	516,37
KP1602	RP941-29-2-1-1-B	45	2	33	68,8	409,1	47,4	0,6	0,9	458,1
KP1603	RP941-29-2-1-2-B	43	2	33	80,7	351,9	46,3	0,4	0,9	399,4
KP1605	RP941-29-2-2-1-	42	2	31	79,4	384,2	35,5	1,3	0,4	421,4

Line ID	Name	Total oil %	Water %	Average oleic acid %	Average linoleic acid %	α-Tocopherol (ppm)	β-Tocopherol (ppm)	γ-Tocopherol (ppm)	δ-Tocopherol (ppm)	Total tocopherol (ppm)
	B									
KP1606	RP951-6-2-4-2-B	43	2	36	84,1	533,9	51,9	1,4	0,9	588,1
KP1608	RP951-6-3-3-2-B	38	2	17	36,8	397,3	48,2	1,2	1	447,8
AGSUN 5264	AGSUN 5264	40	3	17	27,5	264,9	96,9	3,6	2,2	367,6
AGSUN 5217	AGSUN 5217	39	3	19	26,9	171,2	92,6	2,6	2,2	268,6
PAN 7033	PAN 7033	39	3	14	34,4	296,9	98,4	3,5	2,4	401,2
PAN 7049	PAN 7049	39	3	15	63,5	247,9	97,3	3,7	2,4	351,2
PAN 7098	PAN 7098	40	3	16	22,3	236,2	95,2	3,2	2,5	337
PAN	PAN	40	3	16	62,9	283,4	99	4	2,3	388,7

Line ID	Name	Total oil %	Water %	Average oleic acid %	Average linoleic acid %	α-Tocopherol (ppm)	β-Tocopherol (ppm)	γ-Tocopherol (ppm)	δ-Tocopherol (ppm)	Total tocopherol (ppm)
7057	7057									
AGSUN 5278	AGSUN 5278	38	3	18	57,4	307,9	98,6	4,3	2,3	413
SY 4045	SY 4045	39	3	25	49,4	299,4	98,4	4	2,3	404
PAN 7102	PAN 7102	42	2	14	18,1	286,2	97,2	3,3	2,3	389
CAP 4000	CAP 4000	40	3	21	99,4	315,4	100,9	3,9	2,4	422,7
AGSUN 5270	AGSUN 5270	42	2	15	71,5	275,1	96,1	3,7	2,3	377,1
AGSUN 8251	AGSUN 8251	39	3	17	65,3	253,3	96,3	3,2	2,3	355,1
65A70	65A70	39	2	17	79,7	244,4	94,8	3,2	2,2	344,7
AGSUN 5279	AGSUN 5279	25	7	19	48,6	255,1	95,1	3,3	2,3	355,8

Line ID	Name	Total oil %	Water %	Average oleic acid %	Average linoleic acid %	α-Tocopherol (ppm)	β-Tocopherol (ppm)	γ-Tocopherol (ppm)	δ-Tocopherol (ppm)	Total tocopherol (ppm)
PAN 7100	PAN 7100	40	2	13	108,7	255,2	95,1	3,1	2,2	355,6
SY 4200	SY 4200	45	2	18	62	269,2	96	3,7	2,4	371,4
PAN 7080	PAN 7080	41	2	14	60,9	257,2	95,7	2,9	2,3	358,2
PAN 7102	PAN 7102	39	2	22	30,7	235,4	95,2	2,7	2,3	335,5
65A15	65A15	43	2	16	83,3	258,2	94,9	3,3	2,3	358,7
PAN 7095	PAN 7095	38	3	23	53,1	251,3	93,7	2,9	2,2	350,2

[1]Extreme moisture outside calibration range and may influence other values

Chapter 4

Determining vitamin E biosynthetic gene homologues and their genetic diversity within sunflower accessions displaying different tocopherol levels

4.1 Abstract

Linchay Daniels & Dirk (ZH) Swanevelder
Agricultural Research Council, Biotechnology Platform, Onderstepoort, 0110, Pretoria, South Africa
University of South Africa, Florida, 1709, Johannesburg, South Africa
Email: Swanevelderd@arc.agric.za

The oxidative stability and shelf life of sunflower oil is conferred by its tocopherol content. This research aims to: (1) Identify the tocopherol gene homologues and construct a biosynthetic pathway, and (2) to determine if there's any genetic polymorphism (diversity) for these homologues. Vitamin E gene homologues were identified through literature and bioinformatic searches. The vitamin E pathway was constructed, with genes of interest identified for genetic variation with known phenomic variations (chapter 3). A sunflower gene reference database was developed for homogentisate phytyltransferase (*HPT*), *γ*-tocopherol methyltransferase (*TMT*) and tocopherol cyclase (*TC*). Polymorphism was observed for > 90% of the 23 accessions tested which contained 489 SNPs and 145 indels. *TMT* had the highest number of SNPs (424) followed by *HPT* (75), and *TC* (1). However, only 139 SNPs were located in the coding region of these genes. These exon-based genes may play a role in translational changes.

Keywords: sunflower, oxidative stability, tocopherol biosynthetic pathway, *HPT*, *TMT*, *TC*

4.1 Introduction

Sunflower is an important oil seed crop worldwide and is largely used in the production of edible food and animal feed (Gotor et al 2007, Nimet et al 2011). Sunflower oil is mainly characterized by oil saturation levels, e.g. high concentrations of linoleic acid followed by oleic, palmitic and stearic acid (Grompone 2005). High oleic acid containing sunflower seed is rich in tocopherols (vitamin E) and phytosterols, which is highly beneficial for health reasons (Gotor et al 2007). Tocopherols are antioxidants that protect the cell against oxidative damage and prevents or delays off flavour and smell (Hunter and Cahoon 2007). Tocopherols are also commercially used in cosmetics, sunscreens and has the potential value as livestock feed supplements to improve the quality and shelf life of meat (Waylan et al 2002). Sunflower has a higher levels of tocopherol compared to other edible oils (Nimet et al 2011). Tocopherols are the most powerful natural antioxidants and contribute to the nutritional value of the oil (Zilic et al 2010).

The biosynthetic pathway of vitamin E has naturally occurring *α, β, γ,* and *δ* tocopherols and tocotrienols. The major genes that lead to the production of these natural occurring tocopherols are ρ-hydroxyphenylpyruvate dioxygenase (*HPPD*), homogentisate phytyltransferase (*HPT*), 2-methyl-6-prenylbenzoquinol methyltransferase (*MSBQ*), *γ*-tocopherol methyltransferase (*TMT*), and tocopherol/tocotrienol cyclase (*TC*) (figure 4.2, Results). The vitamin E biosynthetic enzymes are found in the chloroplast and chromoplasts of various crops, e.g. potatoes, tomatoes, spinach, etc. (Hunter and Cahoon 2007). Two substrates are needed for vitamin E biosynthesis: The first substrate homogentisate (*HGA*) is derived from p-hydroxyphenylpyruvate (HPP) through a decarboxylation that is catalyzed by hydroxyphenylpyruvate dioxygenase (*HPPD*). *HPP* in turn is derived from tyrosine via the shikimate pathway (Hunter and Cahoon 2007). The second substrate required for vitamin E biosynthesis is phytyldiphosphate (*PDP*) or geranylgeranyldiphosphate (*GGDP*). *GGDP* can be reduced to *PDP* for tocopherol synthesis. These substrates initiate the synthesis of the four

tocopherol derivatives but are further methylated by the enzymes p-hydroxyphenylpyruvate dioxygenase (*HPPD*), homogentisate prenyltransferase (*HPT*), tocopherol cyclase (*TC*) and tocopherol methyltransferase (*TMT*) to produce *α, β, γ* and *δ* tocopherol (Hunter and Cahoon 2007).

The vitamin E pathway has been elucidated for spinach over 30 years ago (Soll and Schultz 1979), with the enzymatic steps involved characterized by Soll in 1987. However recent genetic work on *Arabidopsis* mutants and transgenic plants have made it easier to dissect the vitamin E pathway in more detail (Hunter and Cahoon 2007). Almeida et al (2011) used genetic and genomic tools to identify and clone genes of the vitamin E biosynthetic pathway in tomatoes. The vitamin E pathway consists of various enzymes involved in the methylation of tocopherol derivatives but only the enzymes below the two initial substrates, i.e. *HPT*, *MSBQ*, *TC* and *TMT*, are of relevance since it leads to the production of *α, β, γ* and *δ* tocopherol. *HPPD* was shown to be a cytosolic enzyme, which has an influence on the tocopherol content in tobacco (Falk et al 2005). The gene encoding *HPT* has previously been characterized from *Synechocystis* and *Arabidopsis,* with gene synonyms *slr1736*, *VTE2*, *HPT1*, *At2g18950* (Collakova and DellaPenna 2001, Savidge et al 2002, Schledz et al 2001). The gene *MSBQ* was characterized from *Synechocystis* (*sll0418*) and *Arabidopsis* (*VTE3* and *At3g63410*) with *Arabidopsis* also having a mutant, *vte3-1* (Shintani et al 2002, Van Eenennaam et al 2003). *TC* was characterized from sunflower, *Synechocystis* and *Arabidopsis,* as *slr1737*, *VTE1* and *At4g32770* respectively, with mutant's *vte1*, *vte1-1* and *vte1-2* (Porfirova et al 2002, Sattler et al 2003). Lastly the gene *TMT* was characterized from maize (*SXD1*), potato (*StSXD1*), *Synechocystis* (*slr0089*) and *Arabidopsis* (*VTE4*, *At1g64970*) (Hofius et al 2004, Provencher et al 2001, Shintani and DellaPenna 1998).

Despite all the earlier molecular tools available, which generate high-density maps and contained chromosome sequences and EST libraries, progress for identifying the vitamin E pathway within sunflower has been hampered by

map-based cloning approaches (Lai et al 2005) and the unavailability of a sequenced genome. At the time of initiating this work, the vitamin E pathway gene homologues were not yet known for sunflower, nor the diversity within these genes.

Aims

The first aim of this work was to identify the gene homologues of the vitamin E biosynthetic pathway for sunflower. This was attempted by using known vitamin E genes from other species to determine candidates in a pre-released, un-annotated sunflower draft genome and other public bioinformatic available resources. These candidates were then be used to construct a possible biosynthetic pathway for sunflower that was confirmed after sequencing.

Chapter 3, determined the correlation between tocopherols and unsaturated fatty acids content in selected sunflower accessions, and showed that different sunflower accessions in the ARC's sunflower germplasm collection had different tocopherol levels and compositions. This suggests that there may be some variation in either the expression of the biosynthetic pathway genes or mutations with these genes causing different production levels. The second aim of this chapter was therefore to investigate the genetic diversity of these identified biosynthetic gene homologues (Homogentisate phytyltransferase (*HPT*), ρ-Hydroxyphenylpyruvate dioxygenase (*HPPD*), tocopherol cyclase (*TC*), and tocopherol methyltransferase (*TMT*) in accessions already shown to produce varied levels and compositions of tocopherols (chapter 3) using a next generation sequencing (NGS) approach. The analyses of the genetic diversity of the vitamin E genes in sunflower germplasm accessions would be of great interest for the identification of possible mutants for this important economical trait.

4.2 Materials and methods

4.2.1 Sunflower sample selection and collection

In chapter 3, 104 sunflower accessions were phenotypically characterized for their tocopherol and fatty acid levels. From these 104 sunflower accessions, 23 lines were selected here for homologues diversity analyses based on their phenotypic profiles (table 4.1). The phenotypes selected consisted of different unique combinations of oleic and linoleic acid, and different levels of *α, β, γ* and *δ* tocopherol. Six groups were identified based on possible expression levels and/or mutations that could explain the observed phenotypes. In short group 1 consisted of low *α*, but high *β, γ*, and *δ* tocopherol, suggesting that the *TMT* conversion to *α* tocopherol is lower expressed thus *γ* tocopherol is formed, but the *TMT* conversion to *β* is high. This could be either due to lower activity through *MSBQ/MPBQ* resulting in more activity through *TC* and *TMT* on the *β* and *δ* tocopherol branch of the vitamin E biosynthetic pathway and/or a lower conversion rate of *γ* tocopherol to *α* tocopherol via *TMT* - which may suggest an introduced substrate bias. Group 2 was characterized by low *α* and *γ,* but high *β* and *δ* tocopherol, which suggests that the gene *MSBQ/MPBQ* may be less active (or lower expressed), resulting in more activity through *TC* and *TMT* on the *β* and *δ* tocopherol branch of the vitamin E biosynthetic pathway. The accessions in group 3 had low levels of all tocopherol derivatives suggesting that the initial enzyme in the pathway (i.e. *HPT*) is less active/lower expressed or a problem with substrate feeding into the system. Group 4 was classified by low *α* and *β* and high *γ* and *δ* tocopehorols, thereby suggesting lower expression levels or a less active *TMT* gene. Group 5 consisted of high *α* but low *β, γ* and *δ* tocopherols. Since *α* tocopherol is usually higher than the other tocopherols derivatives, this would indicate normal flux through the system, thus causing an accumulation of *α* tocopherol. The sunflower accessions in group 6 where characterized by low *α, β* and *γ* tocopherol but high *δ* tocopherol, thus suggesting that *MSBQ/MPBQ* and *TMT* may be less active (or inactive).

Table 4.1 Sunflower accessions selected for homologue diversity screening based on their unique tocopherol (vitamin E) and fatty acid phenotypes (chapter 3). Accessions were divided into six groups based on possible gene mutations (or expression level changes) in *TC*, *TMT*, *HPT* and *MSBQ/MPBQ* that would produce the observed phenotypes. Group 1 consisting of accessions with low *α*, but high *β, γ* and *δ* tocopherols; Group 2 had lines with low *α* and *γ* but high *β* and *δ* tocopherols; Group 3 were characterized with low levels of all the tocopherols; Group 4 had accessions with low *α* and *β* but high *γ* and *δ* tocopherols*;* Group 5 consisted of high *α* but low *β, γ* and *δ* tocopherols and Group 6 had low *α, β* and *γ* tocopherols, but high *δ* tocopherol.

Group	Sample	High oleic (>80%)	Mid oleic (50-75%)	High linoleic (>80%)	Mid linoleic (50-75%)	Low alpha (170-390 ppm)	High alpha (400-680 ppm)	Low beta (33-94 ppm)	High beta (95-105 ppm)	Low gamma (0.3-2.9 ppm)	High gamma (3-5 ppm)	Low delta (0.3-1.5 ppm)	High delta (2-2.7 ppm)	Total Tocopherol (ppm)
1	**Kp245**			✓		✓			✓		✓		✓	368
1	**Kp413**			✓		✓			✓		✓		✓	423
1	**Kp714**			✓		✓			✓		✓		✓	415
1	**Kp810**				✓	✓			✓		✓		✓	390
1	**Kp1212**		✓		✓	✓			✓		✓		✓	391
2	**Kp806**			✓		✓			✓	✓			✓	355
2	**Kp729**	✓			✓	✓			✓	✓			✓	323
2	**Kp1219**		✓		✓	✓			✓	✓			✓	351

Group	Sample	High oleic (>80%)	Mid oleic (50-75%)	High linoleic (>80%)	Mid linoleic (50-75%)	Low alpha (170-390 ppm)	High alpha (400-680 ppm)	Low beta (33-94 ppm)	High beta (95-105 ppm)	Low gamma (0.3-2.9 ppm)	High gamma (3-5 ppm)	Low delta (0.3-1.5 ppm)	High delta (2-2.7 ppm)	Total Tocopherol (ppm)
3	Kp1191			✓		✓		✓		✓		✓		420
3	Kp1185			✓		✓		✓		✓		✓		401
4	Kp697			✓		✓		✓			✓		✓	374
5	Kp1180		✓	✓			✓	✓			✓		✓	637
5	Kp1170		✓	✓			✓	✓		✓		✓		317
5	Kp1190		✓	✓		✓		✓		✓		✓		536
5	Kp704				✓		✓	✓		✓		✓		489
5	Kp1596				✓		✓	✓		✓		✓		516
5	Kp734		✓				✓	✓		✓		✓		506
5	Kp1229		✓				✓	✓		✓		✓		450
6	Kp759		✓			✓		✓		✓			✓	298
6	Kp801		✓		✓	✓		✓		✓			✓	330
6	Kp802		✓			✓		✓		✓			✓	330
6	Kp472		✓		✓	✓		✓		✓		✓		268
6	Kp1227	✓			✓	✓		✓		✓			✓	296

4.2.2 DNA extraction and normalization

Plant leaf disks of the 104 sunflower accessions that were phenotyped (chapter 3) were collected from Potchefstroom field trails during summer (February and March 2013) as well as in Kwa-Zulu Natal during winter (May and June 2013). A total of 4 leaf disks per accession were taken from each of the 104 accessions and stored at -80^{o}C for future use. Only the 23 identified accessions (table 4.1) were further processed.

DNA extractions were performed on the leaf disks, with each leaf disk crushed to a fine powder with ceramic beads using a Fastprep FP120 (Bio 101 Savant). The crushed samples were extracted using the DNeasy Plant Mini kit (Qiagen, Whitehead Scientific Ltd) in accordance to the manufacturer's protocol. The DNA quality and concentration was measured fluoromatically on a Qubit 2.0 fluorometer (Life Technologies Corporation, California) and visualized on 1% agarose gel (Tris(hydroxymethyl)aminomethane, acetic acid and Ethylenediaminetetraacetic acid [TAE]) at 100V and 80 Amps for 50 minutes. Extracted DNA was stored for future experiments on the sunflower germplasm at -20^{o}C in 1X TE buffer (Tris(hydroxymethyl)aminomethane [Tris] and Ethylenediaminetetraacetic acid [EDTA]). Sample working aliquots were normalized to 15 ng/µL.

4.2.3 Candidate gene identification

A literature search was conducted to find the vitamin E genes of interest from crops previously studied (*Arabidopsis*, potatoes, maize, tomatoes, *Lativca satuca* and *Brassica napus*) (table 4.1A, appendix 4) (Hass et al 2006, García-moreno et al 2012). The gene homologues for the tocopherol biosynthetic enzymes of interest where also identified via database searches on NCBI, SRS-EMBL, CGP and TIGR, using these known vitamin E gene IDs, DNA and/or protein fragments (table 1A, appendix 4) from the literature and databases searches. These searches gave information on the gene name, synonym, id, accession number, species, function and location of the gene (table 4.1A, appendix 4). BLAST (Altschul et al 1990) and translated BLAST

(tBLASTx), (Altschul et al 1990) searches were performed to obtain gene candidates of interest (table 4.4) in sunflower. Where sunflower DNA or mRNA gene homologues were obtained, they were compared with reference genes from *Arabidopsis* and the other known plant tocopherol gene to confirm that the sunflower candidates were complete. Where incomplete or no sunflower gene homologues were available, the closest relative known from these databases was used as a candidate species to identify homologues from the un-annotated draft sunflower genome (Version: September 2012) which were made available by Prof Loren Riesenberg (University of British Columbia UBC). The BLAST functions on the CLC Genomics Workbench 9.0 (hppt://www.qiagenbioinformatics.com/) were used for the Sunflower draft genome searches. Possible homologues obtained in this fashion were aligned and compared with sunflower and/or closely related species' genes to confirm identity. Only complete alignments with known members of the gene of interest, were considered possible homologues of the gene of interest and used for primer design.

4.2.4 Primer design

Identified candidate gene homologues of different plants were used in multiple sequence alignment comparisons (alignments are presented on the supplementary CD as figures CD4.1 to CD4.5, with the figure titles presented as the filenames) of the genes of interest to identify conserved regions for possible priming sites. The genes tocopherol cyclase (*TC*, DQ229845) and tocopherol methyltransferase (*TMT*, JN991198) were located for sunflower via the searches. Primer sets were designed using the online program PrimerQuest tools from Integrated DNA technologies (IDT) (http://eu.idtdna.com/Primerquest/Home/Index). Default primer selection settings (percentage GC content (%GC) 50-60%; melting temperature (Tm) of between 55°C and 80°C, with and optimal annealing temperature of 55°C; Stability between 5' and 3' ends (5' vs. 3') of more-or-equal-than 1.2 kcal; Matches greater but equal to 3 will be rejected at 3' end; but with greater and equal to 7 adjacent homolog bases; Base runs greater but equal to 4 will also be rejected which contain more-or-equal to 3 dinucleotide repeats) were used.

Multiple sequence alignments of p-hydroxyphenylpyruvate dioxygenase (*HPPD*), homogentisate prenyltransferase (*HPT*) and 2-methyl-6-prenylbenzoquinol methyltransferase (*MSBQ*) (alignments are presented on the supplementary CD as figures CD4.1 to CD4.5, with the figure titles presented as the filenames) of known gene homologues from related species (e.g. *Luctuca* and *Brassica*) and possible sunflower gene homologous, were used to identify conserved regions in the genes. Putative primer sites within the conserved regions of the gene were identified on the CLC Genomics Workbench 9.0 (hppt://www.qiagenbioinformatics.com/) by conducting a Blast analysis of the hits per gene against the draft sunflower genome**.** The forward and reverse primers were designed according to the number and position of hits on the genes to obtain as long as possible gene fragments while remaining in exons/conserved regions. The oligonucleotide length for all primers was set between 20 to 25 base pairs, based on the primer pair matches under the default settings. Primer sequences identified and used in this chapter are given in table 4.2.

Table 4.2 Primer sequences targeting vitamin E gene homologues in sunflower. The forward and reverse primer sequences for hydroxyphenylpyruvate dioxygenase (*HPPD*), Homogentisate prenyltransferase (*HPT*), 2-Methyl-6-prenylbenzoquinol methyltransferase (*MSBQ*), Tocopherol cyclase (*TC*) and Tocopherol methyltransferase (*TMT*) identified and used are presented here (5' to 3'). Several primer sets were designed for *MSBQ* after difficulty with amplification was experienced.

Gene Name	Forward Primer (5'-3')	Reverse Primer (5'-3')	Expected amplification (bp)
HPPD	GCTCTGAGAAGTTCCCTTTC	ATGTGTTGTCGGATGAGCAG	1547
HPT	AGGTTAACAAGCCCTATCTTCC	GCAACAGCATAAGCCACTTC	1747
***MSBQ* set1**	GCAATCCCACTGAACCAATAAC	CCCAAACCACAGGGACTAAA	4285
***MSBQ* set 2**	GGCTGGAAGGTATTGGTTCTT	GCACCATACGGGTCTCTTAATC	4285
***MSBQ* set 3**	GGCATTGGACTGAGGATATGAG	GAGTCGGTTGGACAAGGTATG	4285
TC	CTGCGCAACTTGAATGGAGCTACAG	GATTCCGTCCACATCAACAGGAAGG	5354
TMT	CATCACCACCACCACCGCCAAATTC	GTCCCGGTATAATCAATGCCGGTAG	4126

4.2.5 Candidate gene amplification and confirmation

Polymerase chain reactions (PCRs) were performed using ThermoScientific$^{(TM)}$ DreamTaq PCR Master mix (2X) (Applied Biosystems, Foster City, California, USA) and the identified primers sets (table 4.2) to target each gene candidate of interest. This polymerase mix consists of an optimized buffer, dNTPs and $MgCl_2$. PCR amplification was carried in a 25 µl final volume with a final concentration of 1X Green DreamTaq master mix, 10 µM of forward and reverse primer, 15 ng DNA template and PCR nucleotide free water. Amplification reactions were carried out on a Multigene Labnet PCR system (Labnet International Inc.) or G-storm (VACUTEC) thermocycler. The PCR optimized conditions for each primer set is presented in table 4.3 by taking into account the primers' annealing temperatures and the extension/ elongation times of the different expected products. The gene *TC* did not amplify with normal PCR settings and a two-step PCR was thus used, i.e. a two-step PCR is combining the annealing and extension steps into a single step to get accurate bands at the desired product size. The gene *MSBQ* did not amplify after many attempts of trouble shooting therefore it was excluded from this chapter.

Amplicons were separated on 1% agarose TAE gels ran on a Cleaver Scientific Ltd gel system at 100V (80 Amps) for 50 minutes. The gels were stained with ethidium bromide (0.5 mg/ml) and viewed on an UV imaging system (UVP-Biospectrum AC Imaging system, Ultra-Violet Products Ltd, Cambridge, UK).

Table 4.3 PCR amplification parameters used in amplifying vitamin E gene homologues. All the reactions starts with an initial denaturing step at 94°C for 2 minutes and ends with a final elongation step of 72°C for 10 minutes and hold at 4°C until samples were removed. All cycling conditions presented here consisted of 30 cycles. The gene *TC* was amplified with a two-step PCR and *MSBQ* did not amplify after many attempts of troubleshooting and was therefore excluded from this project.

Gene	Denaturation	Annealing	Extension
Hydroxyphenylpyruvate dioxygenase (*HPPD*)	94°C @ 30 sec	45°C @ 30 sec	72°C @ 1.5 min
Homogentisate prenyltransferase (*HPT*)	94°C @ 30 sec	59.1°C @ 30 sec	72°C @ 2.1 min
Tocopherol cyclase (*TC*)	94°C @ 30 sec		68 °C @ 5 min
Tocopherol methyltransferase (*TMT*)	94°C @ 30 sec	57°C @ 30 sec	72°C @ 4.3 min
2-Methyl-6-prenylbenzoquinol methyltransferase (*MSBQ*)*	94°C @ 30 sec	47 to 80°C @ 30 sec	72 °C @ 5 min

***** No amplification under tested conditions using various primer sets

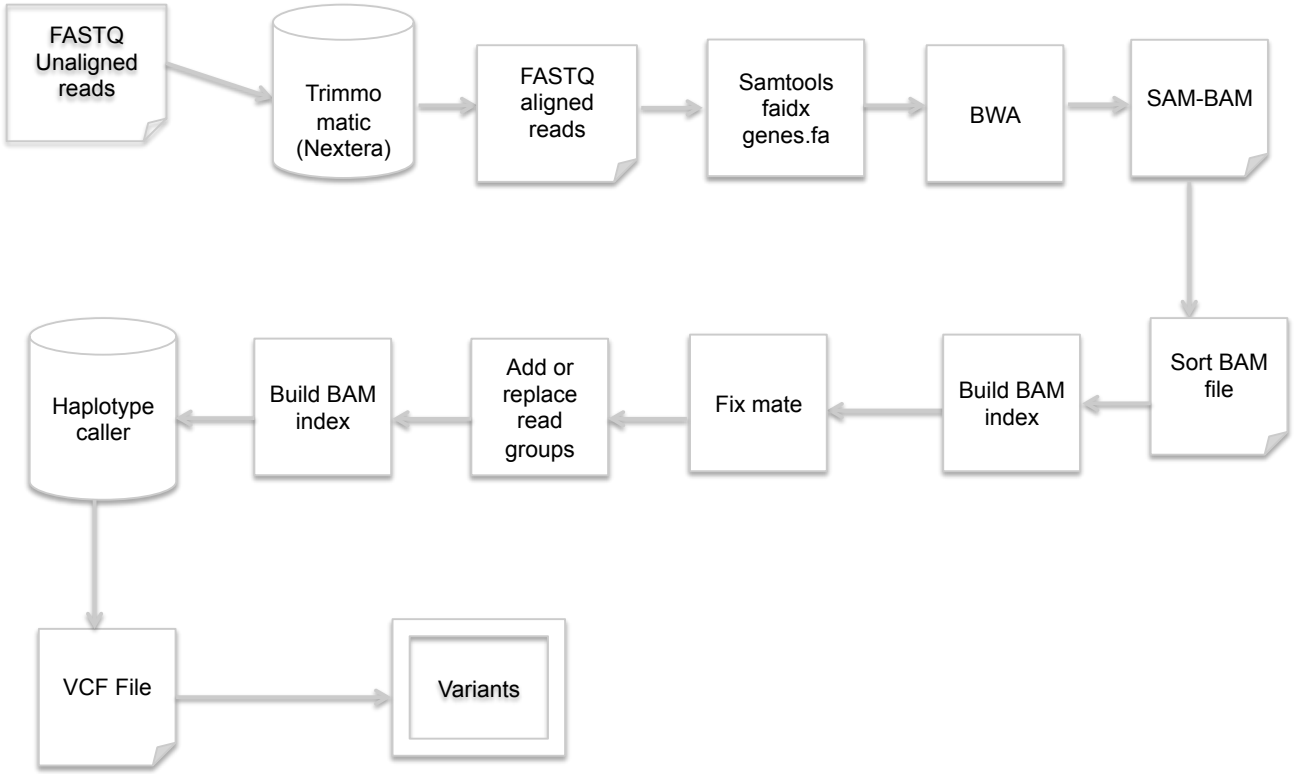
4.2.6 Next generation sequencing (NGS) sample preparation and sequencing

A trail experiment was conducted for the four gene homologues on accession KP 245 and KP 806 to validate if the targeted vitamin E gene homologues were correct (table 4.5, Results) using sequence analyses on CLC Genomics Workbench 9.0 (hppt://www.qiagenbioinformatics.com/) and BLAST-annotation comparison on NCBI and BLAST2GO. Once confirmation of the correct gene homologues was determined, the remainder of the accessions were sequenced and analysed. A pooling strategy was used to lower sample preparation costs. In this strategy, prior to sequencing and sample preparation, all accession's amplicons for the different targets (i.e. the four tocopherol genes) where normalized on concentration and pooled in equal volumes, e.g. amplicons of *HPT, HPPD*, *TC* and *TMT* were pooled for KP 245. Of the 23 sunflower accessions, three accessions had incomplete pooling of the four genes, i.e. KP 704, KP 1219 and KP 1229, with *TC*, *HPPD* and *HPT*, respectively not amplifying and therefore not included within these pools. Individual amplicons were separated for each sample from a pool based on sequence homology. Each pooled amplicon sample was sample prepped for next generation sequencing (NGS). Briefly, the Nextera DNA Library Prep kit (Illumina) was used in which a tagmentase fragments and adds universal NGS adapters to the DNA samples, i.e. the pooled, sample specific PCR products. Unique Nextera indexes (Illumina) were subsequently added by PCR. Libraries were cleaned and sequenced on a MiSeq (Illumina NGS system) using the manufacturer's sequencing protocols and reagents. Pair end (PE) sequencing (300x300 bp sequencing) was performed on the fragments using MiSeq Reagent Kit v3 (Illumina). Samples were loaded to obtain at least 130X average coverage on a standard MiSeq run. The Next Generation Sequencing was performed at the ARC Biotechnology Platform's Sequencing Core Facility.

4.2.7 Data analyses

Sequence analyses were done on the CLC Genomics Workbench 9.0 (hppt://www.qiagenbioinformatics.com/) for the predicted sunflower vitamin E gene homologues that could be consistently and successfully amplified. The reads obtained were filtered, trimmed and aligned using default settings on the CLC Genomics Workbench 9.0 (hppt://www.qiagenbioinformatics.com/) and mapped to the *Helianthus annuus L.* draft reference genome (version 2012). Overlapping pairs were merged to form contigs and a *de novo* assembly was done per accession. Each gene was mapped to the reference gene and consensus sequences were extracted from the mapped reads to perform a BLASTx-annotation and mapping on BLAST2GO. The genes of correct gene ontology (GO) terms, size, name and function were extracted to build a reference database for the four vitamin E genes (*HPT*, *MSBQ*, *TC* and *TC*). Variant calling was done using haplotype caller in Genome Analysis Toolkit (GAKT) (McKenna et al 2010) (figure 4.1). The unaligned reads were trimmed with trimmomatic (Bolger et al 2014) and the aligned reads were viewed with FASTQ (Andrews 2010). The targeted genes were placed in SAMtools (Li et al 2009) and aligned using Burrows-Wheeler Aligner (BWA) (Li and Durbin 2009) to created SAM-BAM files. The SAM and Bam files were sorted to build Bam index files and SNPs were called using Haplotype caller and viewed in Integrative Genomics Viewer (IGV) (Robinson et al 2011). SNPs and indels with a quality score less than 1000 (threshold) were discarded as they were regarded as false positive SNPs.

Figure 4.1 The pre-variant pipeline for the determination of Single Nucleotide Polymorphism (SNP) using the Haplotype caller in GATK. The unaligned reads were trimmed using trimmomatic (Bolger et al 2014). Once all the reads were trimmed and aligned it was then viewed in FASTQ (Andrews 2010). The genes were then inserted in SAMtools (Li et al 2009) where they were aligned using BWA (Li and Durbin 2009) and saved as SAM-BAM files. Haplotype caller was used to call for variants in GATK whereby the output file was saved as a VCF file and variants were viewed using IGV (Robinson et al 2011).



4.3 Results

4.3.1 The biosynthetic pathway of vitamin E in sunflower

The biosynthetic pathway of vitamin E was constructed for sunflower (figure 4.2) by identifying the candidate genes from various crops using literature and different publicly available bioinformatics sites (also see section 4.3.2, Candidate gene identification and primer design). Once possible gene homologues for the vitamin E pathway had been identified in sunflower, *Arabidopsis*, *Lativca satuca* and *Brassica napus*, the candidates were BLAST annotated on BLAST2GO V.2.6.4 (http://www.blast2go.de), mapped with TAIR (http://www.arabidopsis.org/) and remapped in Uniprot (http://www.uniprot.org/). The results were mapped to the KEGG pathways website (http://www.genome.jp/kegg/pathway.html) and a reference pathway: "Ubiquinone and other terpenoid quinone biosynthesis pathway" was obtained (http://www.kegg.jp/kegg-bin/highlight_pathway?scale=1.0&map=map00130&keyword=Ubiquinone%20and%20other%20terpenoid%20quinone%20biosynthesis%20pathway). The biosynthetic pathway of vitamin E was then constructed based on the reference pathway from KEGG (figure 4.2).

.

4.3.2 Candidate gene identification and primer design

Literature results indicated that the genes tocopherol cyclase (*TC*) and tocopherol methyltransferase (*TMT*) were in sunflower, but homogentisate prenyltransferase (*HPT*) and hydroxyphenylpyruvate dioxygenase (*HPPD*) were not found in this species. Gene homologues from different crops were subsequently used to obtain these two genes of interest in sunflower. Different gene homologues were found for *HPT* and *HPPD*. BLAST searches were completed for these on NCBI to get the highest matched homologues for the possible gene candidates (table 4.4). The gene homologues found for *HPT* and *HPPD* were from *Brassica napus,* the closest relative to sunflower compared to other plant species. These genes were not available for sunflower but gene homologues JN 834019 (*HPT*) and JN 834016 (*HPPD*) were used for primer design (table 4.1A, appendix 4) and later annotated against the draft sunflower genome (version 2012) to identify homologues in sunflower. *MSBQ* (DQ229835), *TC* (DQ229849) and *TMT* (JN 991198) were retrieved from sunflower (table 4.1A, appendix 4).

The genomic DNA sequences of these gene homologues were used to create gene maps to observe the position of the exons and introns (figure 4.2). Primers were design based on the whole gene region for all targeted vitamin E genes (table 4.3 and table 4.4), with the exon regions targeted for primer design since these regions are more conserved then the introns e.g. tocopherol/tocotrienol cyclase (*TC*) (5' CTGCGCAACTTGAATGGAGCTACAG 3') and tocopherol/tocotrienol methyltransferase (*TMT*) (5' CATCACCACCACCACCGCCAAATTC 3') using Primer design3 (Rozen and Skaletsky 2000). The fragment length after primer design for *TC* was 5109 bp and 4002 bp for *TMT* respectively. The gene *HPT* consists of 1747 bp but the whole region could not be used for primers design since no stable hybridization site was found at the 5' end at either of the first 5 exons. The forward primer for *HPT* was created on the 6th exon and the reverse primer on the 11th exon. The primers designed for *HPPD* (1547 bp) and *MSBQ* (4285 bp) were based on the whole gene region. Three different sets of primers were designed for *MSBQ* since amplification was difficult to

achieve. Primer sets for all five genes were tested and optimized whereby various PCR parameters were adapted (e.g. annealing temperature, cycle time, etc.) to produce positive results with the correct band size.

Table 4.4 NCBI BLAST homologues for vitamin E gene candidates. Top hits of BLAST homologues for hydroxyphenylpyruvate dioxygenase (*HPPD*), Homogentisate prenyltransferase (*HPT*), 2-Methyl-6-prenylbenzoquinol methyltransferase (*MSBQ*), Tocopherol cyclase (*TC*) and Tocopherol methyltransferase (*TMT*) from publicly available DNA and protein databases (SwissProt and NCBI) are presented here. The genes containing DNA sequences (◆) were selected from the NCBI BLAST top hits (and the remaining mRNA, RNA and transcribe RNA gene sequences were discarded) and further analysed by doing a tBLASTx search against the sunflower draft genome (version 2012). Phylogenetic distances to sunflower were used to select optimal candidate gene sequences for multiple sequence alignment. Candidate genes (■) were obtained and used for primer designing. The resulting sequences were further annotated by doing a BLASTn search on NCBI.

Gene	Gene id and species	NCBI BLAST homologues (RNA, DNA, mRNA)	Gene DNA sequences	BLAST matches against the sunflower draft genome	NCBI BLAST hits (BLASTn)	E-value & % identity	Species	Gene type or description
HPT	gi|186501445 *Arabidopsis thaliana*	gb|JN834019◆ gb|AY063893.1 gi|17104827 gi|42467434 gi|42468088 gb|GU198365.1 gi|21281071 gi|20384918 gi|299507805 gb|FJ362602.1 gb|JF326244.1 gb|BT146336.1	gb|JN834019 (*Brassica napus*)	gnl|BL_ORD_ID|289474 & gnl|BL_ORD_ID|483	DQ423115■	3.93E-12 84.03%	*Helianthus annuus*	At1g16080 (chloroplast protein)

Gene	Gene id and species	gi|242093813 NCBI BLAST homologues (RNA, DNA, mRNA)	Gene DNA sequences	BLAST matches against the sunflower draft genome	NCBI BLAST hits (BLASTn)	E-value & % identity	Species	Gene type or description
					GU198365■	9.57E-10 92.73%	*Arabidopsis thaliana*	homogentisate phytyltransferase (*HPT*)
					NM_179653■	9.57E-10 92.73%	*Arabidopsis thaliana*	homogentisate phytyltransferase (*HPT*)
					AY113993■	9.57E-10 92.73%	*Arabidopsis thaliana*	*At2g18950*
					AY063893■	9.57E-10 92.73%	*Arabidopsis thaliana*	*At2g18950*
					AF324344	9.57E-10 92.73%	*Arabidopsis thaliana*	tocopherol polyprenyltransferase (TPT1)
					BX819447■	9.57E-10 92.73%	*Arabidopsis thaliana*	*At2g18950*
					BX819968■	9.57E-10 92.73%	*Arabidopsis thaliana*	*At2g18950*

Gene	Gene id and species	NCBI BLAST homologues (RNA, DNA, mRNA)	Gene DNA sequences	BLAST matches against the sunflower draft genome	NCBI BLAST hits (BLASTn)	E-value & % identity	Species	Gene type or description
					CP002685	2.35E-10 92.86%	*Arabidopsis thaliana*	*AT2G01008*
					JN834019 ■	3.51E-3 87.50%	*Brassica napus*	homogentisate phytyltransferase BnaX.*VTE2.b*
HPPD	gi|238478368 *Arabidopsis thaliana*	gi|238478368 gb|AF060481.1 gi|145335185 gb|AC007592.4 gb|AY072329.1 gb|AF047834.1 gb|CP002684.1 gb|JN834026.1 gb|DQ886526.1 gb|JN834016.1 ◆ gi|312282468	gb|JN834016.1 (*Brassica napus*)	gnl|BL_ORD_ID|289038 & gnl|BL_ORD_ID|47 & gnl|BL_ORD_ID|313759 & gnl|BL_ORD_ID|24768	XM_002520323	8.96E-80 86.75%	*Ricinus communis*	RCOM_1192240
					XM_002300831	7.71E-68 85.49%	*Populus trichocarpa*	POPTRDRAFT_798588
					AB267400	2.72E-58 81.94%	*Coptis japonica*	*hppd*

Gene	Gene id and species	NCBI BLAST homologues (RNA, DNA, mRNA)	Gene DNA sequences	BLAST matches against the sunflower draft genome	NCBI BLAST hits (BLASTn)	E-value & % identity	Species	Gene type or description
					AJ309203	1.08E-57 81.84%	*Coleus blumei*	*hppd*_SOLSC
					XM_004144581	6.40E-53 83.19%	PREDICTED: *Cucumis sativus*	LOC101215051 4hydroxyphenylpyruvate dioxygenase-like
					AY957391	1.56E-50 81.15%	*Medicago truncatula*	HPD
					AY957391	1.56E-50 81.15%	*Medicago truncatula*	4hydroxyphenylpyruvate dioxygenase (HPD)
					AM423697	5.51E-41 80.18%	*Vitis vinifera*	VITISV_033517 VITISV_033518
					XM_004243561	8.60E-40 82.49%	*PREDICTED: Solanum lycopersicum*	LOC101245475

Gene	Gene id and species	NCBI BLAST homologues (RNA, DNA, mRNA)	Gene DNA sequences	BLAST matches against the sunflower draft genome	NCBI BLAST hits (BLASTn)	E-value & % identity	Species	Gene type or description
					XM_004303840	1.34E-38 88.68%	PREDICTED: *Fragaria vesca subsp. vesca*	LOC101313724 4hydroxyphenylpyruvate dioxygenase-like
					JN834016■	8.28E-37 81.82%	*Brassica napus*	4hydroxyphenylpyruvate dioxygenase
					XM_003617336	1.29E-35 82.71%	*Medicago truncatula*	MTR_5g090990 4hydroxyphenylpyruvate dioxygenase
					AB376089	2.02E-34 81.67%	*Hevea brasiliensis*	*hppd*
					JN834026■	7.98E-34 81.11%	*Brassica napus*	4hydroxyphenylpyruvate dioxygenase (*BnaX.PDS1.a*)
					XM_004491260	1.25E-32 80.71%	PREDICTED: *Cicer arietinum*	LOC101503806

Gene	Gene id and species	NCBI BLAST homologues (RNA, DNA, mRNA)	Gene DNA sequences	BLAST matches against the sunflower draft genome	NCBI BLAST hits (BLASTn)	E-value & % identity	Species	Gene type or description
					AY138969	4.92E-32 79.44%	*Abutilon theophrasti*	*HPPD*
					XM_004240123	7.68E-31 81.38%	PREDICTED: *Solanum lycopersicm*	LOC101257377
					AJ309203	3.04E-30 87.94%	*Coleus blumei*	*hppd*_SOLSC
					JN834015■	4.74E-29 80.50%	*Brassica napus*	4-hydroxyphenylpyruvate dioxygenase (*BnaX.PDS1.b*)
					XM_002300831	1.87E-28 86.08%	*Populus trichocarpa*	POPTRDRAFT_798588
					NM_001248219	2.93E-27 80.82%	*Glycine max*	LOC100101901

Gene	Gene id and species	NCBI BLAST homologues (RNA, DNA, mRNA)	Gene DNA sequences	BLAST matches against the sunflower draft genome	NCBI BLAST hits (BLASTn)	E-value & % identity	Species	Gene type or description
					XM_002307526	2.82E-24 90.91%	*Populus trichocarpa*	POPTRDRAFT_559505
					AY957391	1.67E-19 83.65%	*Medicago truncatula*	4hydroxyphenylpyruvate dioxygenase (*HPPD*)
					NM_001160839	9.95E-15 83.45%	*Arabidopsis thaliana*	*PDS1* (gene synonym 4-hydroxyphenylpyruvate dioxygenase)
					GQ860303	8.96E-66 80.54%	*Sesamum indicum*	Chl P

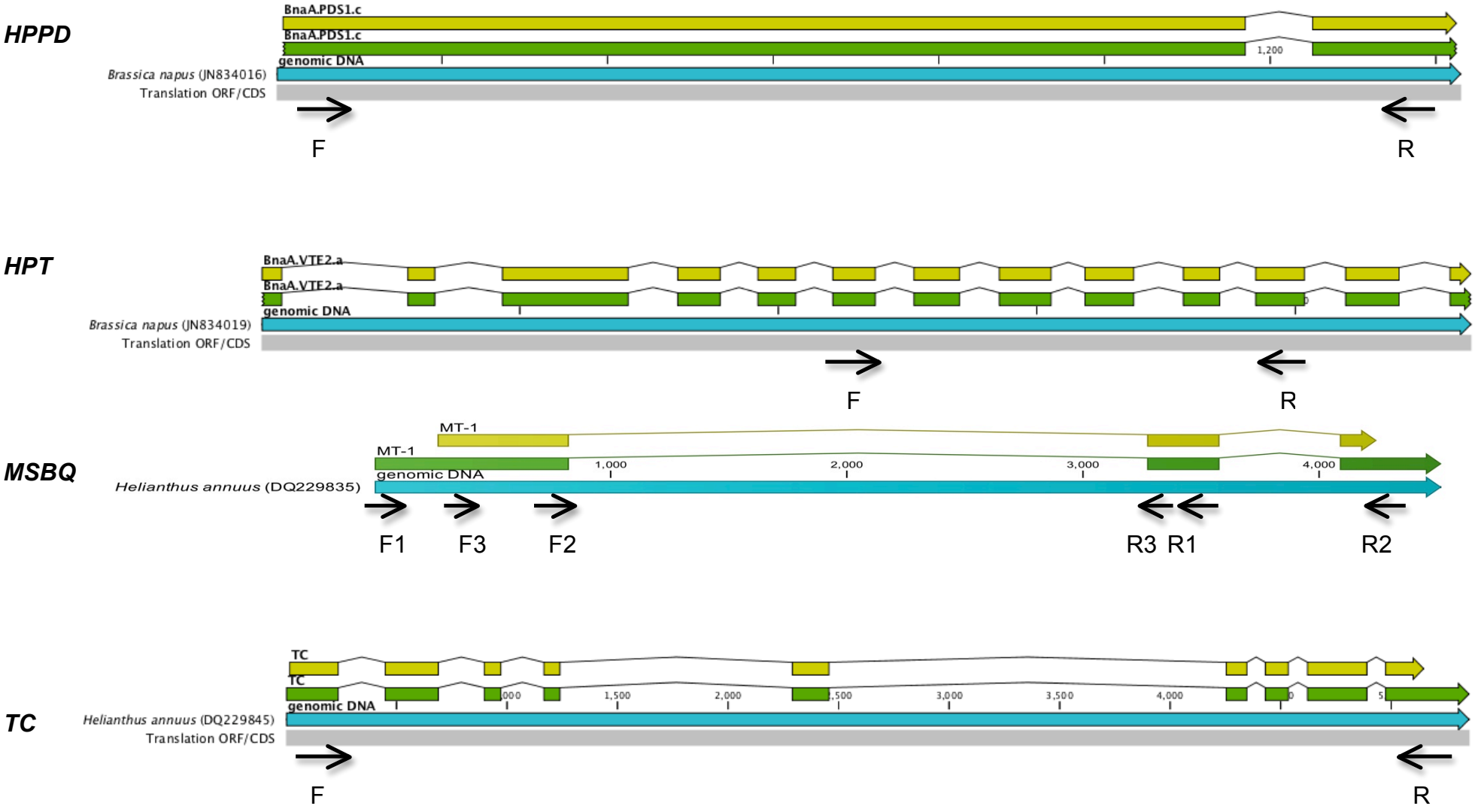
Gene	Gene id and species	NCBI BLAST homologues (RNA, DNA, mRNA)	Gene DNA sequences	BLAST matches against the sunflower draft genome	NCBI BLAST hits (BLASTn)	E-value & % identity	Species	Gene type or description
MSBQ	gi|80971663 *Helianthus annuus*	gi|80971663◆ gi|80971665 gb|DQ229837.1 gi|80971679 gb|DQ229838.1 gb|DQ229843.1 gb|DQ229841.1 gb|DQ229840.1 gb|DQ229842.1	gi|80971663| DQ229835 *(Helianthus annuus)*	gnl|BL_ORD_ID|296826 & gnl|BL_ORD_ID|7835	DQ229835■	*4.03E-17 100%*	*Helianthus annuus*	*MPBQ/MSBQ* methyltransferase 1
					DQ229836■	1.20E-177 96.45%	*Helianthus annuus*	*MPBQ/MSBQ*
					DQ229838■	1.50E-158 99.00%	*Helianthus annuus*	*MT-1*
					DQ229838■	1.50E-158 99.00%	*Helianthus annuus*	*MT-1*
					DQ229839■	1.50E-158 99%	*Helianthus annuus*	*MT-1*
					DQ229837■	1.19E-66 95.68%	*Helianthus annuus*	*MPBQ/MSBQ*

Gene	Gene id and species	NCBI BLAST homologues (RNA, DNA, mRNA)	Gene DNA sequences	BLAST matches against the sunflower draft genome	NCBI BLAST hits (BLASTn)	E-value & % identity	Species	Gene type or description
					XM_004167357	1.10E-60 83.71%	*PREDICTED: Cucumis sativus*	LOC101227374
					DQ229844■	2.40E-49 81.89%	*Helianthus annuus*	*MT-2*
					DQ229843■	2.40E-49 81.89%	*Helianthus annuus*	*MPBQ/MSBQ*
					DQ229842■	2.40E-49 81.89%	*Helianthus annuus*	*MPBQ/MSBQ*
					DQ229841■	2.40E-49 81.89%	*Helianthus annuus*	*MPBQ/MSBQ*
					DQ229840■	2.40E-49 81.89%	*Helianthus annuus*	*MPBQ/MSBQ*
					XM_003590510	5.86E-47 81.84%	*Medicago truncatula*	MTR_1g071110
					FJ269356■	1.28E-35 96.94%	*Helianthus annuus*	FAD2-1 RPA1
					AB514532	7.59E-31 80.16%	*Diospyros kaki*	Dk671

Gene	Gene id and species	NCBI BLAST homologues (RNA, DNA, mRNA)	Gene DNA sequences	BLAST matches against the sunflower draft genome	NCBI BLAST hits (BLASTn)	E-value & % identity	Species	Gene type or description
					JF490078	4.51E-26 79.71%	*Arachis hypogaea*	*MPBQ/MSBQ*
TC	gi|80971681 *Helianthus annuus*	gi|80971681◆ gi|80971689 gi|80971687 gi|80971685 gb|DQ229846.1	gi|80971681| gb|DQ229845.1 *(Helianthus annuus)*	gnl|BL_ORD_ID|295086 & gnl|BL_ORD_ID|6095	DQ229845■	1.15E-37 100%	*Helianthus annuus*	tocopherol cyclase (*TC*)
					DQ229849■	1.02E-37 100%	*Helianthus annuus*	tocopherol cyclase (*TC*)
					DQ229848■	1.02E-37 100%	*Helianthus annuus*	tocopherol cyclase (*TC*)
					XM_003590510	5.86E-47 81.84%	*Medicago truncatula*	MTR_1g071110
					DQ229847■	1.02E-37 100%	*Helianthus annuus*	tocopherol cyclase (*TC*)
					DQ229846■	1.02E-37 100%	*Helianthus annuus*	tocopherol cyclase (*TC*)

Gene	Gene id and species	NCBI BLAST homologues (RNA, DNA, mRNA)	Gene DNA sequences	BLAST matches against the sunflower draft genome	NCBI BLAST hits (BLASTn)	E-value & % identity	Species	Gene type or description
gamma-*TMT*	gi|371766612 *Helianthus annuus*	gi|371766616 gb|DQ229832.1 gb|DQ229831.1 gb|DQ229828.1 gb|DQ229829.1 gb|JN991201.1 gb|JN991198.1 ◆ gb|DQ229833.1	gi|371766612 | gb|JN991198 (*Helianthus annuus)*	gnl|BL_ORD_ID|29004	DQ229828■	*6.58E-54 100%*	*Helianthus annuus*	gamma-*TMT*
					JN991201■	*4.39E-95 95.52%*	*Helianthus annuus*	gamma-*TMT*
					DQ229834■	*2.61E-90 94.62%*	*Helianthus annuus*	*gamma-TMT*
					JN991201■	*4.39E-95 95.52%*	*Helianthus annuus*	gamma-*TMT*
					DQ229834■	*2.61E-90 94.62%*	*Helianthus annuus*	*gamma*-TMT
					DQ229833■	*2.61E-90 94.62%*	*Helianthus annuus*	*gamma-TMT*
					JN991201.1 ■	*9.22E-81 91.44%*	*Helianthus annuus*	*gamma-TMT*
					JN991199■	2.32E-44 96.71%	*Helianthus annuus*	*gamma-TMT*

Gene	Gene id and species	NCBI BLAST homologues (RNA, DNA, mRNA)	Gene DNA sequences	BLAST matches against the sunflower draft genome	NCBI BLAST hits (BLASTn)	E-value & % identity	Species	Gene type or description
				gnl|BL_ORD_ID|1050	DQ229828■	*1.23E-52 100%*	*Helianthus annuus*	gamma-*TMT*
				gnl|BL_ORD_ID|318507	JN991201.2 ■	*6.00E-130 90.88%*	*Helianthus annuus*	gamma-*TMT*


HPPD
BnaA.PDS1.c
BnaA.PDS1.c
genomic DNA
Brassica napus (JN834016)
Translation ORF/CDS
1,200
F
R
HPT
BnaA.VTE2.a
BnaA.VTE2.a
genomic DNA
Brassica napus (JN834019)
Translation ORF/CDS
F
R
MSBQ
MT-1
MT-1
genomic DNA
Helianthus annuus (DQ229835)
1,000
2,000
3,000
4,000
F1
F3
F2
R3
R1
R2
TC
TC
genomic DNA
Helianthus annuus (DQ229845)
Translation ORF/CDS
1,500
2,000
3,000
3,500
4,000
F
R

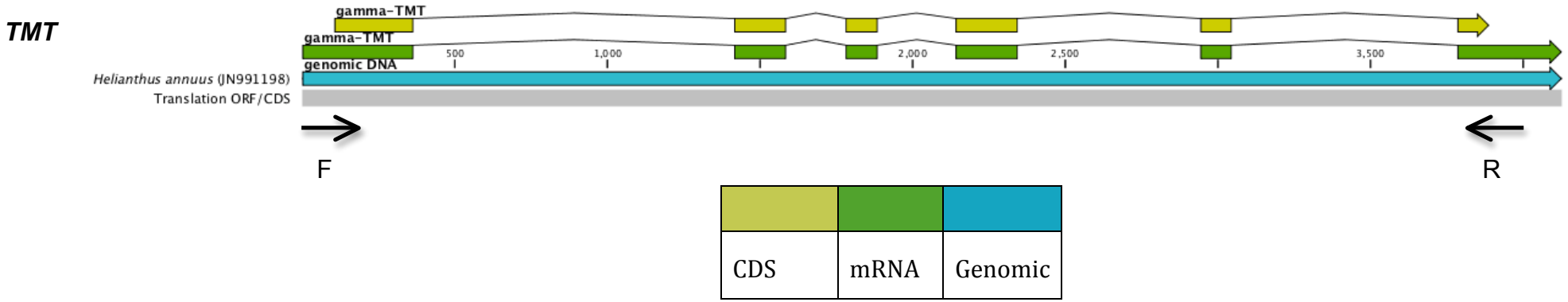


Figure 4.2 Gene maps of vitamin E homologues in sunflower and other closely related species indicating open reading frames, introns and exons. The vitamin E gene maps of hydroxyphenylpyruvate dioxygenase (*HPPD*), homogentisate prenyltransferase (*HPT*), 2-methyl-6-prenylbenzoquinol methyltransferase (*MSBQ*), tocopherol cyclase (*TC*) and tocopherol methyltransferase (*TMT*) were constructed using genomic DNA and mRNA accessions from Genbank (https://www.ncbi.nlm.nih.gov). *Brassica napus* (JN834016 and JN834019) were used for genes that are not yet annotated in sunflower (*HPPD* and HPT respectively). The positions of the introns and exons ensure primers were designed in conserved exon regions of the genes. The different colour strains indicate the coding strain (CDS) (yellow), mRNA (green) and genomic DNA (blue) for each gene. Blocks indicate codons and introns connect exons with lines. The sequences were downloaded in CLC Genomics Workbench 9.0 (hppt://www.qiagenbioinformatics.com/) (http://www.clcbio.com) for the construction of each gene map. The primer sites designed for an arrow indicates each construct, with F indicating the forward and R the reverse primer.

4.3.3 PCR amplification

PCR amplification using the specific primer combination for each gene (table 4.2) produced amplicon sizes of about 1000 bp for *HPPD*, 2000 bp for *HPT*, 5000 bp for *TC* and 4000 bp for *TMT* (figure 4.3). For *MSBQ* several primer sets were tested at various temperature gradients and touchdowns, but amplification continuously showed multiple bands. After several attempts on different sunflower accessions this gene still failed to amplify at the correct size (4126 bp) and consequently was therefore removed from this project. The gene *TMT* amplified at the expected size (4126 bp) for all sunflower accessions analysed. Standard PCR settings were initially tested for *TC* but variable amplifications were observed or nothing at all. A two-step PCR was done and produced the expected band at approximately 5354 bp (table 4.2 and 4.3).

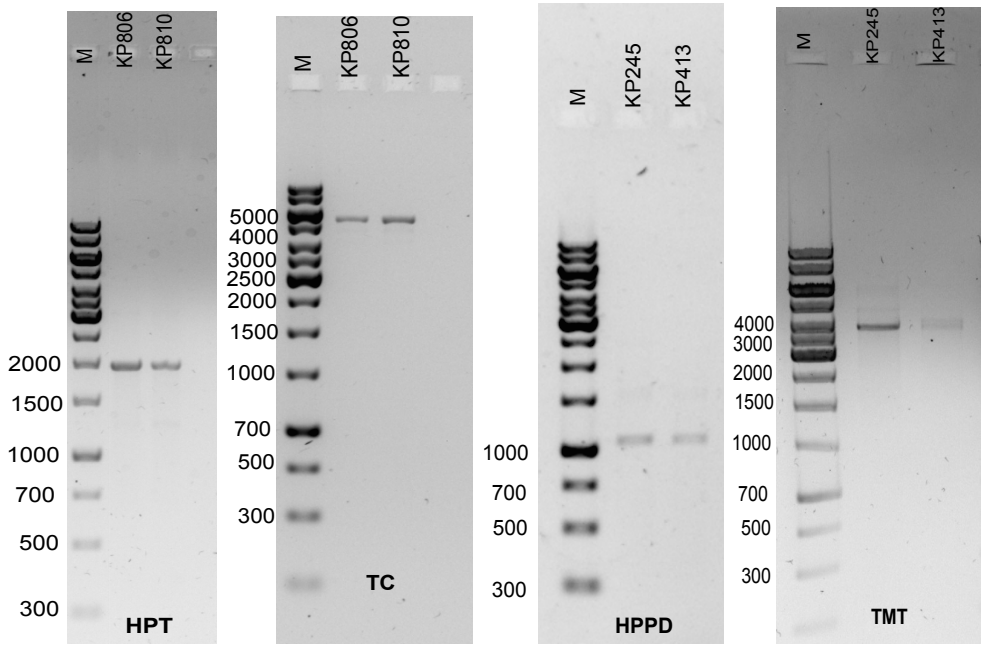


Figure 4.3 PCR amplicons indicating product sizes. The expected gene homologues of *HPT* (KP 806 and KP 810), *TC* (KP 806 and KP 810), *HPPD* (KP 245 and KP 413) and *TMT* (KP 245 and KP 413). The gene *HPT* amplified at 2000 bp, *TC* at about 5000 bp, *HPPD* at 1000 bp and *TMT* at ±4126 bp using the 1Kb (300-10000 bp) marker.

PCR amplification for *HPPD* occurred at 1000 bp as suppose to 1500 bp for all 23 accessions but was still sequenced to observe why this gene amplified at this specific size. BLAST annotation indicated that the amplification at 1000 bp is homogentisate phytylprenyltransferase (*HPPT*) rather than *HPPD*. According to KEGG *HPPT* is part of the tocopherol/tocotrienol biosynthesis. This gene can lead to the production of either tocopherol or tocotrienol via homogentisate phytyltransferase (*HPT*) or homogentisate geranylgeranyl transferase (*HGGT*) depending on the type of antioxidant produced. Very little literature is available for *HPPT* but it has been characterized in *Glycine max* (soybean) from accession Q58FG4, with gene ID 732648 (Uniprot, NCBI and KEGG). It is believed that this gene is involved in the transferase activity transferring alkyl or aryl groups but not methyl groups. Sequencing results for the 23 accessions were used to construct a sunflower gene database for *HPT*, *HPPT*, *TC* and *TMT*.

4.3.4 Gene validation and reference gene database

Sequence analyses were done for a trail experiment using a subset of the targeted amplicons and analysed on the CLC Genomics Workbench 9.0 (hppt://www.qiagenbioinformatics.com/) for the four targeted vitamin E gene homologues. The reads obtained were filtered, trimmed and aligned to the *Helianthus annuus* draft reference genome (version 2012) and overlapping pairs were merged to form contigs. The contigs were BLAST-annotated by comparison to the NCBI nucleotide database and then compared to sunflower and/or closely related species to confirm and identify the amplicons (table 4.5). Following the trail experiment testing the amplicons for the genes of interest, the remaining accessions were sequenced and analysed. A “reference sunflower database” was generated from the resultanting sequence analyses and BLAST annotations for hydroxyphenylpyruvate dioxygenase (*HPPD*), homogentisate prenyltransferase (*HPT*), tocopherol cyclase (*TC*) and tocopherol methyltransferase (*TMT*). This was accomplished after the sequence analysis on CLC Genomics Workbench 9.0 (hppt://www.qiagenbioinformatics.com/) by extracting the consensus sequence from the mapped reads, followed by a BLASTx, annotation and

mapping on BLAST2GO. The genes with the correct sequence description, size, gene ontology and highest E-value were selected to build the sunflower gene reference database.

Table 4.5 Next generation sequencing (NGS) analysis of a trail experiment to validate the amplicons as the correct gene homologues when targeting vitamin E genes in sunflower. The validation of homogentisate phytyltransferase (*HPT*), ρ-hydroxyphenylpyruvate dioxygenase (*HPPD*), tocopherol cyclase (*TC*), and tocopherol methyltransferase (*TMT*) was confirmed through next generation sequencing with the reads trimmed, merged and mapped to the reference genes using CLC Genomics Workbench 9.0 (hppt://www.qiagenbioinformatics.com/). The consensus sequences were extracted and a nucleotide BLAST was done on NCBI to confirm the gene identity.

Gene	Species from which primers were designed	Gene ID	Gene size (bp)	Read count	Matched reads	Unmatched reads	Percentage identity matched to NCBI (Genbank)
HPPD	*Lactuca savita*	225 001 451	1743	1 244 184	1 156 560	87 624	80
HPT	*Brassica napus*	377 657 554	1670	356 930	157 166	199 764	91
TC	*Helianthus annuus*	80 971 685	5354	711 884	658 054	53 830	100
TMT	*Helianthus annuus*	80 971 654	4035	616 750	514 991	101 759	96

4.3.5 Identification of polymorphism within the vitamin E gene amplicons

In this study 23 sunflower accessions were categorized in six groups according to its phenotypic characteristics (table 4.1). These groups represent possible mutations or regulation differences in the genes involved in the biosynthetic pathway as predicted from the phenotypes characterization in chapter 3. SNPs and indels were identified in 16 of the 23 sunflower accessions tested (table 4.6). The seven accessions: KP 245, KP 714, KP802, KP 810, KP 1212, KP 1180 and KP 1229 lacked DNA polymorphisms in the regions compared in this study. The sequencing results revealed 489 total SNPs and 145 total indels across the three targeted vitamin E genes homogentisate prenyltransferase (*HPT*), tocopherol cyclase (*TC*) and tocopherol methyltransferase (*TMT*) (table 4.6). The fourth gene hydroxyphenylpyruvate dioxygenase (*HPPD*) is not included for the SNP analyses because, sequencing results revealed that *HPPD* was not present within the accessions analysed but rather homogentisate phytylprenyltransferase (*HPPT*). *HPPT* was not further analysed since only 747 bp was recovered after sequencing analysis and 397 bp had no sequence coverage. Therefore this gene gave false positive SNPs over gap regions where there was no sequence coverage available, thus this gene was removed from this project.

The highest number of total SNPs (197) was found for group one with 165 SNPs for the gene *TMT* and 32 for *HPT*. From the 165 *TMT* SNPs only 44 were found on the exon and 121 on the intron. The accession with highest number of SNPs in group one is KP 806 with 92 SNPs but only 26 of these SNPs fall within the exons, followed by KP 413 with 69 SNPs where 18 were found in the exons (table 4.6). Group one also had the highest number of SNPs (32) and indels (7) for the gene *HPT* but 12 of these SNPs where found on the exon and 20 on the introns. The group that represents the second highest total number of SNPs (134) is group five with 45 indels. The 134 SNPs that were found are for *TMT* (131), *HPT* (2) and *TC* (1). From the 131 *TMT* SNPs only 38 are found on the exons (24 for KP 1170 and 14 for KP

1596). The single allele found for *TC* comes from the accessions KP 734 but this SNP was found on an intron and a single deletion was found on exon 6 at position 4258. This gene was only found once across all the accessions analysed with a SNP at position 4597 and were conserved for five of the six phenotypically characterized groups (table 4.2A, appendix 4). The exon regions code for proteins whereas introns are not implicated with protein coding. Therefore all SNPs that were found on the coding region might induce amino acid code changes for proteins, which in turn may cause functional changes in the gene structure.

Group two represents the second highest number of SNPs for *HPT* with 29 total SNPs where 15 of these SNPs were found on the exons. The gene *TMT* had the highest number of SNPs with a total number of 424 SNPs across the six phenotypically characterized groups, where 112 were found on the exons and 312 on the introns. This is followed by *HPT* with 64 SNPs in total where 27 were found on the exons and 37 on the introns. The lowest number of SNPs (two) and indels (zero) were found in group four (KP 697) for the gene *HPT* but both of these SNPs were found on the introns.

The results also revealed that several sunflower accessions had a SNP for *TMT* at position 3832. These accessions are KP 413, KP 472, KP 729, KP 806, KP 1170, KP 1190, KP 1219 and KP 1596. The SNP that occurred at position 3832 was found on the exon six for all eight accessions mentioned above. An insertion where G was replaced with GTCTT was also observed on exon two at position 1440 for some accessions in group one, three, five and six. These accessions that share a SNP, insertion or deletion at the same position across several accessions might be of importance because its found on the exon of the gene and can cause functional changes, but need further investigation.

A total of 145 indels were found across the sunflower accessions analysed. The highest number of indels was found in group one with 52 indels for *TMT* (where 28 are insertions and 24 deletions) and seven for *HPT* with six insertions and one deletion. Group five had the second highest number of

indels (45), with 44 for *TMT* (which consist of 21 insertions and 23 deletions) and one deletion for *TC* at position 4285. Group four is the only group that had no indels present as expected because this group only had one accession present.

Some indels were found in several accessions at the same position e.g. *TMT* had an insertion at position 1102 for KP 806, KP 1185, KP 1227 and KP 1596. Insertions and deletions were also found for *TMT* at position 1106 for KP 729, KP 759 and KP 1170 but at both positions (1102 and 1106) these insertions and deletions were found on intron regions. Some indels was identified as insertions or deletions for *TMT* e.g. KP 806, KP 1190 and KP 1596 had a deletion at position 2241 where CATAGTG was replaced with a C (table 4.2A, appendix 4). An insertion for *TMT* was observed at position 1786 for KP 413, KP 806 and KP 1596 where A was replaced with ATAG. In both instances these insertions and deletions were observed on exon three (position 1786) and exon four (position 2241). Thus since these insertions and deletions was found on the exon regions of *TMT* further investigations need to be done to determine if these insertions or deletions cause any changes in the function of the proteins.

Table 4.6 SNP identification for homogentisate prenyltransferase (*HPT*), tocopherol cyclase (*TC*) and tocopherol methyltransferase (*TMT*). A total of 23 accessions were analysed for SNPs and indels and these were found in 16 of the 23 accessions. The remaining 7 accessions (KP 245, KP 714, KP802, KP 810, KP 1212, KP 1180 and KP 1229) had no SNPs (-) and indels present. Group one had the highest total number of SNPs and indels present while group 4 had the lowest total number of SNPs and indels.

Group	Sample	SNPs (*TMT*)	Indel (*TMT*)	SNPs (*HPT*)	Indel (*HPT*)	SNPs (*TC*)	Indel (TC)
1	KP 245	-	-	-	-	-	-
Low *α*; High *β, γ & δ*	KP 413	69	25	-	-	-	-
	KP 472	2	-	-	-	-	-
	KP 714	-	-	-	-	-	-
	KP 729	2	1	32	7	-	-
	KP 806	92	26	-	-	-	-
	Total	**165**	**52**	**32**	**7**	**-**	**-**
2	KP 810	-	-	-	-	-	-
(Low *α* & *γ*; High *β* & *δ*)	KP 1212	-	-	-	-	-	-
	KP 1219	2	-	29	6	-	-
	Total	**2**	**-**	**29**	**6**	**-**	**-**
3 (Low *α*, *β*, *γ*; & *δ*)	KP1185	12	1	-	-	-	-

	KP 1190	48	12	-	-	-	-
	KP 1191	10	1	-	-	-	-
	Total	**70**	**14**	**-**	**-**	**-**	**-**
4 (Low *α* & *β*; High *γ* & *δ*)	KP 697	-	-	2	-	-	-
	Total	**-**	**-**	**2**	**-**	**-**	**-**
5 (High *α*; Low *β*, *γ* & *δ*)	KP 1180	-	-	-	-	-	-
	KP 704	-	-	2	-	-	-
	KP 734	-	-	-	-	1	1
	KP 1170	74	18	-	-	-	-
	KP 1229	-	-	-	-	-	-
	KP 1596	57	26	-	-	-	-
	Total	**131**	**44**	**2**	**-**	**1**	**1**
6 (Low *α*, *β* & *y*; High *δ*)	KP 759	-	1	-	-	-	-
	KP 801	-	-	-	1	-	-
	KP 802	-	-	-	-	-	-
	KP 1227	55	19	-	-	-	-
	Total	**55**	**20**	**-**	**1**	**-**	**-**

The SNP distributions for the 23 sunflower accessions were measured as the total number of SNPs per accession across all genes (%). Seven of the 23 accessions had no SNPs identified in the gene regions compared. The highest SNP distribution was found for KP 806 (19%), across the 23 accessions analysed. This is followed by the accession KP 1170 with a 15% SNP distribution, KP 413 with 14%, and KP1569 with 12% (figure 4.4). The accession KP 1227 had a SNP distribution of 11%, KP 1190 10% and KP 729 with 7%. The rest of the sunflower accessions had a SNP distribution between 0 and 6% (figure 4.4).

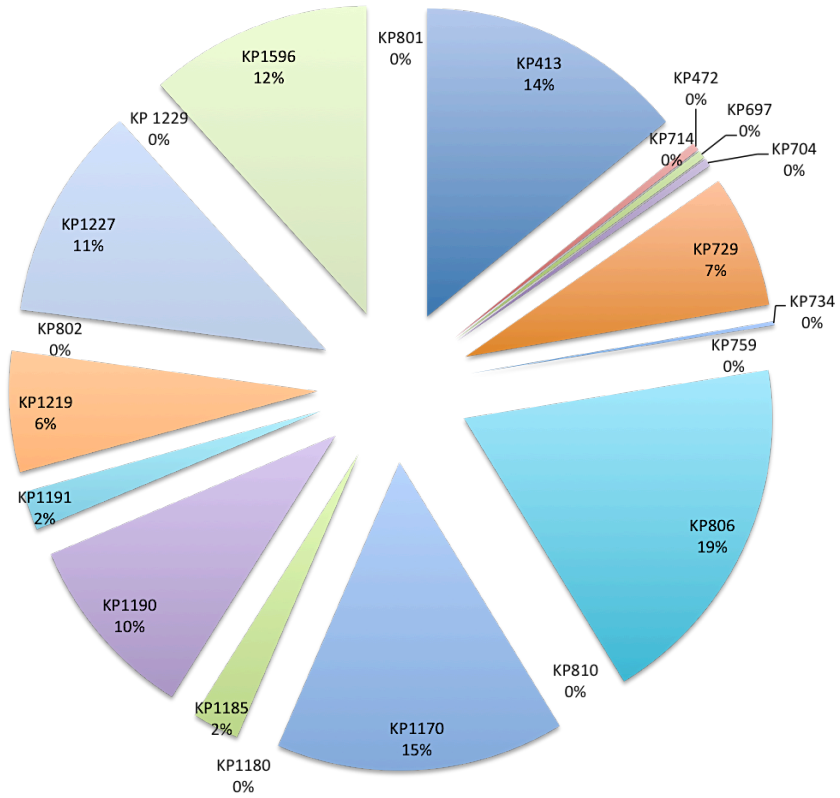


Figure 4.4 SNP distributions across the 23 sunflower accessions. The highest SNP distribution was found for KP 806 based on the total number of SNPs found per accession for all genes analysed. There are several accessions present with no SNPs namely KP 245, KP 714, KP 802, KP 810, KP 1212, KP 1180 and KP 1229.

4.4 Discussions

Soll et al (1980) have elucidated the vitamin E biosynthetic pathway several years ago. The genes encoding the enzymes of this pathway have been characterized and isolated in *Arabidopsis thaliana* and *Synechocystis* (Porfirova et al 2002, Collakova and DellaPenna 2003, Van Enennaam et al 2003). In this study the gene homologues of interest were successfully identified, isolated and characterized from publicly available resources. The resulting contigs were BLAST-annotated in a comparison to sunflower and closely related species to confirm the identity of the genes. The gene IDs were used to construct a vitamin E biosynthetic pathway based on a similar reference pathway found on KEGG. Custom designed primers for homogentisate prenyltransferase (*HPT*), hydroxyphenylpyruvate dioxygenase (*HPPD*), tocopherol cyclase (*TC*) and tocopherol methyltransferase (*TMT*) were successfully designed and used for amplification. Hydroxyphenylpyruvate dioxygenase (*HPPD*) amplification was problematic with a 1000 bp fragment being amplified instead of a 1500 bp fragment. Even though this gene amplified at 1000 bp for all 23 sunflower accessions, it was still sequenced to uncover if this is the gene of interest or a closely related relative. Sequencing analysis of the "*HPPD* fragments" and a BLASTx-annotation and mapping search revealed that this gene was not *HPPD*, but rather a close family member called homogentisate phytylprenyltransferase (*HPPT*). Further searches confirmed that this gene is involved in the production of tocopherol or tocotrienols. Little literature was found for this gene but several bioinformatics sites (Uniprot, KEGG and NCBI) indicated that *HPPT* was characterized in *Glycine max*. Depending on the type of vitamin E synthesized *HPPT* can either help in the production of tocopherol via homogentisate phytyltransferase (*HPT*) or tocotrienol via homogentisate geranylgeranyl transferase (*HGGT*) (Hunter and Cahoon 2007). The gene *HPPT* was present in some of the 23 accessions, but only a portion of the gene (747 bp) was recovered after sequencing analysis. The 747 bps had a huge gap of 397 bp (350 bp – 747 bp) calling false positive SNPs (with a quality score below 1000) where there was no sequence coverage. Therefore

the SNP analysis for this gene was not included in this project. SNP discovery was successfully done for three (*HPT*, *TC* and *TMT*) out of the four gene homologues analysed. From the sequencing results a sunflower gene database was successfully build for *HPT*, *TMT* and *TC*.

The six phenotypically characterized groups (chapter 3) illustrate the relation between the four tocopherol derivatives and the genes involved. Group one had high *β, γ* and *δ* but low *α* indicating that the *TMT* conversion to *α* tocopherol is low but that of *β* tocopherol is high. According to Cheng et al (2003) and Van Enenennaam et al (2003) the methylation of *TMT* disrupts the synthesis of *α* and *β* tocopherol and causes a build up of *γ* and *δ* tocopherol. This indicates that the methylation of the enzyme *MSBQ/MPBQ* is less or inactive, suggesting that the enzymes *HPT*, *TC* and *TMT* catalyze the methylation of high *β, γ* and *δ* tocopherol derivatives. A total of 165 *TMT* and 32 *HPT* SNPs have been found for this specific group. From the 165 *TMT* SNPs, 121 were found in the non-coding region (introns) and 44 on the coding region (exons). *TMT* had the largest percentage introns (70%) over the region amplified, which contribute to the biggest part of the gene being non-coding (figure 4.2). Therefore more SNPs were expected on the intron regions of *TMT* compared to the other gene regions. Only 12 of the 32 *HPT* SNPs were found on the exon region because *HPT* consist of 54% introns and 46% exons (figure 4.2). The intron region is less conserved indicating that their sequences can frequently change over time and are removed through RNA splicing to form mature RNA molecules. Although the introns do not code for genes, they may play a significant role in the gene expression level through the formation of secondary and tertiary structures of DNA, RNA and proteins, transcription initiation, elongation and premature transcription termination (Chorev and Carmel 2012). The exon of a gene contains sequences that code the protein product of the gene and are more conserved than the intron, to retain the function of the encoding protein. Therefore since the 44 *TMT* and 12 *HPT* SNPs were found on the exon region, these SNPs might cause changes in the protein structure and therefore influence its function.

The sunflower accessions in group two were characterized by low *α* and *γ* but high *β* and *δ*, suggesting that the gene *MSBQ* may be inactive or less active since the methylation of these genes from *γ* tocopherol to *α* tocopherol is low. Cheng et al (2003) and Van Enenennaam et al (2003) describes how the *MSBQ/MPBQ* gene in *Synechocystis* and *Arabidopsis* disrupts the flow of *γ* and *α* tocopherol on the biosynthetic pathway of vitamin E, redirecting the flow through the *MSBQ/MPBQ*, to the *δ* and *β* tocopherol branch. The gene *TMT* catalyzes the methylation of *β* tocopherol from *δ* tocopherol on the biosynthetic pathway of vitamin E. Only two SNPs were found for *TMT* but both of these SNPs fall in the non-coding region of the gene. A total of 29 *HPT* SNPs were found where only 15 were found on the coding region. The methylation from *γ* to *α* was low, suggesting that there's more activity on the *δ* and *β* branch of the vitamin E pathway since these two tocopherol derivatives were present in higher levels. Therefore, it's suggested that *MSBQ* or an upstream mutation may have redirected the flow from *γ* and *α* tocopherol to the *β* and *δ* tocopherol branch resulting in high levels of *β* and *δ* tocopherol.

Group three contained accessions with low *α, β, γ* and *δ* tocopherol, suggesting that *HPT*, or upstream of it, may be less active since all tocopherol derivatives following this route are low. The first step of vitamin E biosynthesis is the transfer of a prenyl group to homogentisate, which can occur via *HPT* (tocopherol synthesis) or HGGT (tocotrienol synthesis) (Hunter and Cahoon 2007). The substrate used determines whether the final tocochromanol is a tocopherol (*HPT*) or tocotrienol (HGGT) (Hunter and Cahoon 2007). Hunter and Cahoon (2007) found that the overexpression of *HPT* in *Arabidopsis* leaves causes an increase in tocopherol levels as does *HPPD* when overexpressed in transgenic *Arabidopsis* lines. The activities of *HPT* and *HPPD* limited the tocopherol accumulation to some degree in *Arabidopsis* leaves (Hunter and Cahoon 2007). Therefore, since tocopherol production is low for all four derivatives, an upstream regulating mechanism may have hampered the tocopherol flow in the sunflower accessions analysed. In this group a total of 70 SNPs were found for *TMT*, where 11 of these SNPs were found on the exon region of the gene. No SNPs or indels were observed for

the genes *HPT* and *TC* but the 11 SNPs found for *TMT* may cause functional changes in the gene structure.

The accessions for group four consisted of low *α* and *β* but high *γ* and *δ* tocopherol, suggesting that *TMT* might be less active since the end products (*α* and *β*) are low. Shintani and Della Penna (1998) indicated that the methylation of *δ* and *γ* tocopherol leads to the production of *β* and *α* tocopherol but a disruption can be caused in the synthesis of *β* and *α* tocopherol by a *γ-TMT* mutation. This mutation will cause a build up of *δ* and *γ* tocopherol with proportions of the latter hinging on the activity of *MSBQ/MPBQ*. The gene homologue *TMT* appears to be less active on *α* and *β* tocopherol branch since there's no indication of any mutations that could have occurred. The results indicate that no SNPs or indels was found for *TMT* and *TC*, whereas two SNPs were found for the *HPT* gene homologue but both these SNPs (*HPT*) were found on the non-coding region of the gene.

High levels of *α* tocopherol and low levels of *β*, *γ* and *δ* tocopherol were found for group five, suggesting that all enzymes are active and enhancing the production of *α* tocopherol. Literature indicates that sunflower consists of more than 90% *α* tocopherol when one or more upstream methyltransferase activities are disrupted by mutations making the diversity of the tocopherol derivatives unparalleled (Tang et al 2006). According to Cheng et al (2003) and Van Eenennaam et al (2003) the genes *MSBQ/MPBQ* and *TMT* are needed for the synthesis of *α* tocopherol. Therefore since *α* tocopherol was present in levels between 171 – 698 ppm it's suggested that *TMT* was highly active in the methylation process of *γ* to *α* tocopherol. Various SNPs have been found for this group with 131 for *TMT*, 2 for *HPT* and 1 for *TC* indicating that all genes where active in the production of *α* tocopherol. From the 131 *TMT* SNPs only 38 were found on the coding region of the gene and none of the *HPT* and *TC* SNPs were found on the exon region. This indicates that possible mutations (38 *TMT* – SNPs on the coding region) may have disrupted and influenced the methyltransferase activity increasing the diversity of *α* tocopherol.

Group six consisted of low *α, β* and *γ* but high levels of *δ* tocopherol suggesting that the methylation between *MSBQ/MPBQ* and *TMT* from *γ* to *α* tocopherol might be less active. No variation was found for *HPT* and *TC*, but 55 SNPs were found for *TMT*. In total 20 of the 55 SNPs were found on the coding region. The accession KP 1227 was the only accession that contained SNPs (20 on the exon and 35 on the intron). Therefore, since the methylation of *γ* to *α* is low but *δ* tocopherol is high it's suggested that the 20 SNPs found on the coding region of *TMT* have caused a mutation that stopped the production at *δ* tocopherol, resulting in *TMT* on the *γ* and *α* branch to be less active compared to the *δ* branch on the biosynthetic pathway of vitamin E.

4.5 Conclusions

Knowledge of SNPs diversity for the vitamin E genes and if it has any impact on the production of tocopherol derivatives could be a valuable aid in crop improvement strategies for this trait. SNPs can be used to determine expression patterns and possible mutations that can be targeted to breed for specific traits of interest. In this study the gene homologues involved in the production of tocopherol were successfully identified, isolated and characterized using the sunflower draft genome (version 2012) and public bioinformatic available resources. This information was used to find a reference pathway on KEGG, which was used to construct the vitamin E biosynthetic pathway for sunflower. The gene homologues were amplified with custom designed primers and sequenced for SNP determination. The results obtained from sequencing data were successfully used to create a sunflower reference gene database for *HPT*, *TMT* and *TC*.

A total of 489 SNPs and 145 indels were found across the 23 sunflower accessions. Various numbers of SNPs were discovered within each of the six groups classified based on the phenotypic characteristics and the genes involved in the production of vitamin E. Group one had the highest number of SNPs (197) (where 57 were in the coding region) and group four contained the least (2) amount of SNPs (and also the smallest sample number) but none of the 2 SNPs were found in the coding region of the gene. Group six had no variation for almost all the sunflower accessions except for KP 1227 which consisted of 55 SNPs of which 20 were found on the exon region of the gene. The gene *TMT*, the gene with the largest intron percentage, had the highest number of SNPs (424) and was found in five of the six phenotypically characterized groups. A total of 121 from the 424 SNPs were found on the coding region for this gene homologue (*TMT*).

Tocopherol cyclase (*TC*) was conserved for > 95% of the sunflower accession except for KP 734 where a single allele was found. This allele was found on the non-coding region of the gene. The gene *HPPD* was not found but *HPPT*

was found instead, a gene involved in the production of tocopherol or tocotrienol derivatives. No further analysis was done for this gene since only a fragment of the gene (747 bp) was recovered after sequence analysis with a consisted of a gap of 397 bps. This gene picked up false positive SNPs in the regions with no sequence coverage and consisted of a quality score below 1000 (threshold). All SNPs that was found on the coding region of the genes may have a functional since exons are known to code for proteins and are very conserved. In future markers can be developed for the traits of interest once further investigations are made for the SNPs and indels found on the exon regions of the genes or on those on the introns that may introduce stop codons, thus confirming if any functional mutations occurred. It is therefore recommended that further investigations are made for the gene *MSBQ/MPBQ* since literature confirms that *MSBQ/MPBQ* is important in the biosynthetic pathway of vitamin E indicating that it's responsible for the production of *α* tocopherol in physically matured sunflower seeds (Shintani and Della Penna 1998, Cheng et al 2003 and Van Eenennaam et al 2003).

References

Almeida J, Quadrana L, Asís R, Setta N, de Godoy F, Bermúdez L, Otaiza SN, Corrêa da Silva JV, Fernie AR, Carrari F and Rossi M 2011. Genetic dissection of vitamin E biosynthesis in tomato. *Journal of Experimental Botany* 1-18

Altschul SF, Gish W, Miller W, Myers E and Lipman DJ 1990. Basic local alignment search tool. *Journal of Molecular Biology* 403-410

Andrews S 2010. FastQC: A quality control tool for high throughput sequence data. Available online at: hppt://www.bioinformatics.babraham.ac.uk/projects/fastqc

Bolger AM, Lohse M and Usadel B 2014. Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics* 170

Cheng Z, Sattler S, Maeda H, Sakuragi Y, Bryant DA, DellaPenna D 2003. Highly divergent methyltransferases catalyze a conserved reaction in tocopherol and plastoquinone synthesis in cyanobacteria and photosynthetic eukaryotes. *Plant Cell* 2343-2356

Collakova E and DellaPenna D 2001. Isolation and functional analysis of homogentisate phytyltransferase from *Synechocystis* sp. PCC 6803 and *Arabidopsis*. *Plant Physiology* 1113-1124

Falk J, Brosch M, Scha¨fer A, Braun S and Krupinska K 2005. Characterization of transplastomic tobacco plants with a plastid localized barley 4-hydroxyphenylpyruvate dioxygenase. *Journal of Plant Physiology* 738-742

García-moreno MJ, Fernández-martínez JM, Velasco L and Pérez-vich B 2012. Genetic basis of unstable expression of high gamma-tocopherol content in sunflower seeds. *BioMed Central Plant Biology* 1-14

Gotor AA, Berger M, Farkas E, Labalette F, Centis S and Calmon A 2007. Quantification of sunflower minor components by near infrared spectrometry (NIRS). *Helia* 183-190

Grompone M 2005. Sunflower oil. *Bailey's Industrial Oil and Fat Products* 655-730

Hajjar R, Jarvis DI and Gemmill-Herren B 2008. The utility of crop genetic diversity in maintaining ecosystem services. *Agriculture Ecosystems and Environment* 261-270

Harrison I, Laverty M and Sterling E 2004. Genetic diversity. *Openstax-CNX* 1-6

Hass CG, Tang S, Leonard S, Traber MG, Miller JF and Knapp SJ 2006. Three non-allelic epistatically interacting methyltransferase mutations produce novel tocopherol (vitamin E) profiles in sunflower. *Theoretical and Applied Genetics* 767-782

Hofius D, Hajirezaei MR, Geiger M, Tschiersch H, Melzer M and Sonnewald U 2004. RNAi-mediated tocopherol deficiency impairs photoassimilate export in transgenic potato plants. *Plant Physiology* 1256-1268

Hughes AR and Stachowicz JJ 2004. Genetic diversity enhances the resistance of a sea grass ecosystem to disturbance. *Proceeding of the National Academy of Sciences USA* 8998-9002

Hunter SC and Cahoon EB 2007. Enhancing vitamin E in oilseeds: Unraveling tocopherol and tocotrienol biosynthesis. *Lipids* 97-108

Kamal-Eldin A and Appleqvist LA 1996. The Chemistry and Antioxidant Properties of Tocopherols and Tocotrienols. *Lipids* 1-31

Lai Z, Livingstone K, Zou Y, Church SA, Knapp SJ, Andrews J and Rieseberg LH 2005. Identification and mapping of SNPs from ESTs in sunflower. *Theoretical and Applied Genetics* 1532-1544

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R and 1000 Genome Project Data Prossecing Subgroup 2009. The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 2078-2079

Li JT, Yang J, Chen DC, Zhang XL and Tang ZS 2011. An optimized mini-preparation method to obtain high-quality genomic DNA from mature leaves of sunflower. *Genetics and Molecular Research* 1-8

Li H and Durbin R 2010. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* 1754-1760

Mandel JR, Dechaine JM, Marek LF and Burke JM 2011. Genetic diversity and population structure in cultivated sunflower and a comparison to its wild progenitor, *Helianthus annuus L*. *Theoretical and applied genetics* 693-704

Martin GB 1998. Gene discovery for crop improvement. *Current Opinion of Biotechnology* 220-6

Marroni F, Pinosio S, Di Centa E, Jurman I, Boerjan W, Felice N, Cattonaro F and Morgante M 2011. Large-scale detection of rare variants via pooled multiplexed next-generation sequencing: towards next-generation Ecotilling. *The Plant Journal* 736-745

Nimet G, da Silva EA, Palú F, Dariva C, Freitas LDS, Neto AM and Filho LC 2011. Extraction of sunflower (*Heliantus annuus L.*) oil with supercritical CO_2

and subcritical propane: Experimental and modeling. *Chemical Engineering Journal* 262-268

Porfirova S, Bergmuller E, Tropf S, Lemke R and Dormann P 2002. Isolation of an *Arabidopsis* mutant lacking vitamin E and identification of a cyclase essential for all tocopherol biosynthesis. *Proceeding of the National Academy of Science USA* 12495-12500

Provencher LM, Miao L, Sinha N and Lucas WJ 2001. Sucrose export defective1 encodes a novel protein implicated in chloroplast-to-nucleus signaling. *Plant Cell* 1127-1141

Rao R and Hodgkin T 2002. Genetic diversity and conservation and utilization of plant genetic resources. *Plant Cell, Tissue and Organ Culture* 1-19

Rauf S, Teixeira da Silva JA, Khan AA and Naveed A 2010. Consequences of plant breeding on genetic diversity. *International Journal of Plant Breeding* 1-21

Rippert P, Scimemi C, Dubald M, Matringe M 2004. Engineering plant shikimate pathway for production of tocotrienol and improving herbicide resistance. *Plant Physiology* 92-100

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G and Mesirov JP 2011. Integrative Genomics Viewer. *Nature Biotechnology* 24-26

Rozen S and Skaletsky H 2000. Primer3 on the www for general users and for biologist programmers. In: Krawetz S and Misener S, Bioinformatics methods and protocols: Methods in molecular biology. *Humana Press, Totowa* 365-386

Sattler SE, Cahoon EB, Coughlan SJ, DellaPenna D 2003. Characterization of tocopherol cyclases from higher plants and cyanobacteria. Evolutionary implications for tocopherol synthesis and function. *Plant Physiology* 2184-

2195

Savidge B, Weiss JD, Wong YH, Lassner MW, Mitsky TA, Shewmaker CK, Post-Beittenmiller D and Valentin HE 2002. Isolation and characterization of homogentisate phytyltransferase genes from *Synechocystis* sp. PCC 6803 and *Arabidopsis*. *Plant Physiology* 321-332

Schledz M, Seidler A, Beyer P and Neuhaus G 2001. A novel phytyltransferase from *Synechocystis* Sp. PCC 6803 involved in tocopherol biosynthesis. *FEBS Letters* 15-20

Shintani and DellaPenna 1998. Elevating the vitamin E content of plants through metabolic engineering. *Science* 2098-2100

Shintani DK, Cheng Z and DellaPenna D 2002. The role of 2-methyl-6-phytylbenzoquinone methyltransferase in determining tocopherol composition in *Synechocystis* sp. PCC6803. *FEBS Letters* 1-5

Soll J and Schultz G 1979. Comparison of geranylgeranyl and phytyl substituted methylquinols in the tocopherol synthesis of spinach chloroplasts. *Biochemical and Biophysical Research Communications* 715-720

Soll J 1987. α-Tocopherol and plastoquinone synthesis in chloroplast membranes. *Methods in Enzymology* 383-392

Sujatha HL, Chikkadevaiah C and Nandini R 2002. Assessment of genetic diversity among 51 inbred sunflower lines. *Helia* 101-108

Tajima F1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 585-595

Tang S, Hass CG and Knapp SJ 2006. Ty3/gypsy-like retrotransposon knockout of a 2-methyl-6-phytyl-1,4-benzoquinone methyltransferase is non-

lethal, uncovers a cryptic paralogous mutation, and produces novel tocopherol (vitamin E) profiles in sunflower. *Theoretical and Applied Genetics* 783-799

Van Eenennaam AL, Lincoln K, Durrett TP, Valentin HE, Shewmaker CK, Thorne GM, Jiang J, Baszis SR, Levering CK, Aasen ED, Hao M, Stein JC, Norris SR, Last RL 2003. Engineering vitamin E content: from Arabidopsis mutant to soy oil. *Plant Cell* 3007-3019

Waylan AT, O'Quinn PR, Unruh JA, Nelssen JL, Goodband RD, Woodworth JC, Tokach MD and Koo SI 2002. Effects of modified tall oil and vitamin E on growth performance, carcass characteristics, and meat quality of growing-finishing pigs. *Journal of Animal Science* 1575-1585

Zilic S, Maksimovic-Dragisic J, Maksimovic V, Maksimovic M, Basic Z, Crevar M and Stankovic G 2010. The content of antioxidants in sunflower seed and kernel. *Helia* 75-84

Appendix 4 (chapter 4)

Table 4.1A Tocopherol gene information. Gene information for all four vitamin E genes that were retrieved from publicly available databases (e.g. NCBI) and articles

Gene name	Gene synonym and/ ID	Genbank accession number	Species previously isolated from	Location	Functions	Reference articles
ρ-Hydroxyphenylpyruvate dioxygenase (*HPPD*)/ hydroxyphenylpyruvate dioxygenase (*PDS1*)	4-Hydroxyphenylpyruvate dioxygenase *F12K11.9* *F12K11_9* *HPD* *PDS1* phytoene desaturation 1	NM_001160839 (*Arabidopsis thaliana*) JN 834016 (*Brassica napus*)	*Arabidopsis* *Lativa satuca* (Garden lettuce) *Brassica napus*	Cytosol	- *4HPPD* catalyzes the formation of homogentisate (2,5-dihydroxyphenylacetate) from 4HPP and molecular oxygen.	Ren et al 2011 Hunter and Cahoon 2007 Garcia et al 1999
Homogentisate phytyltransferase (*HPT*)/ homogentisate phytyltransferase 1 (*HPT1*)	*slr1736* *VTE2* *HPT1* *At2g18950* *ATHPT;* *F19F24.15;* *F19F24_15;* TOCOPHEROL PHYTYLTRANSF	NM_179653 (*Arabidopsis thaliana*) JN 834019 (*Brassica napus*)	*Synechocystis* *Arabidopsis* *Zea mays* *Brassica napus* *Lactuca sativa* (Garden lettuce)	Chloroplast and chromoplast	Encodes homogentisate phytyltransferase involved in tocopherol biosynthesis. Has impact on seed longevity and plays a role in the adaptation to low temperature	Maeda et al 2008 Hunter and Cahoon 2007 Ren et al 2011

Gene name	Gene synonym and/ ID	Genbank accession number	Species previously isolated from	Location	Functions	Reference articles
	ERASE*;* TOCOPHEROL POLYPRENYLT RANSFERASE*; TPT1; VITAMIN E 2; VTE2*				stress, notably phloem loading.	
Prenylbenzoquinol methyltransferase (PrBQMT) / 2-methyl-6-phytylbenzoquinone methyltransferase/ MPBQ methyltransferase	Albino or pale green mutant1 APG1 E37 IEP37 INNER ENVELOPE PROTEIN 37 Vitamin E defective 3 *VTE3*	DQ229835 (*Helianthus annuus*) DQ229839 (*Helianthus annuus*)	*Helianthus annuus* *Arabidopsis* *Brassica napus*	Chloroplast (Inner envelope membrane)	Encodes 2- methyl-6-phytyl-1,4-benzoquinone/2-methyl-6-solanyl- 1,4-benzoquinone methyltransferase (*MPBQ/MSBQ- MT*) in the inner envelope membrane.	Hass et al 2006 Motohashi et al 2003
Tocopherol/tocotrienol cyclase (*TC*)	*slr1737* *VTE1* *At4g32770*	DQ229849 (*Helianthus annuus*)	Sunflower *Synechocystis* *Arabidopsis*	Chloroplast	Tocopherol cyclase limits tocopherol synthesis in leaves, and the simultaneous loss of tocopherol and glutathione affects photosynthesis.	Hass et al 2006 Kanwischer et al 2005

Gene name	Gene synonym and/ ID	Genbank accession number	Species previously isolated from	Location	Functions	Reference articles
Tocopherol/tocotrienols *γ*-methyltransferase (*TMT*)	*SXD1* *StSXD1* *slr0089* *VTE4* *At1g64970*	JN991198 (*Helianthus annuus*)	Sunflower Maize Potato *Synechocystis* *Arabidopsis*	Chloroplast (Inner membrane)	*TMT* catalyzes the final step in synthesis of *α* or *β* tocochromanols: the methylation of the number 5 carbon on the chromanol ring.	García-moreno et al 2012 Hunter and Cahoon 2007

Table 4.2A Single nucleotide polymorphism (SNP) for all 23 samples for the 3 genes analysed (*HPT*, *TC* and *TMT*). Each variant is characterized by the position of the SNP or indel, the reference allele, alternative allele, frequency and quality score. SNPs and indels with a quality score less than 1000 (threshold) was discarded for both because it was regarded as false positives.

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
KP 413 70 SNP & 23 Indels	*TMT*	222	C	T	66	34	SNP	3081.77
	TMT	229	T	C	65	35	SNP	3115.77
	TMT	231	A	G	65	35	SNP	3082.77
	TMT	264	C	T	60	40	SNP	2675.77
	TMT	444	T	TA			Indel	1581.73
	TMT	445	T	A	64	36	SNP	3275.77
	TMT	454	A	G	58	42	SNP	3676.73
	TMT	466	AC	A			Indel	3676.73
	TMT	515	T	G	62	38	SNP	3342.77
	TMT	528	TGAAACTCAAAG CAAAA	T			Indel	3179.77
	TMT	559	C	T	72	28	SNP	2894.77
	TMT	568	G	A	69	31	SNP	26577
	TMT	598	G	C	64	36	SNP	3076.77
	TMT	632	T	A	65	35	SNP	3051.77
	TMT	658	T	C	64	36	SNP	2423.77
	TMT	700	C	CGTGTTTAAT			Indel	1301.73
	TMT	702	C	CATAATTATTAA			Indel	1437.73
	TMT	704	T	TCGGGAACC			Indel	2077.73

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
	TMT	717	TA	T			Indel	1954.73
	TMT	779	A	G	79	21	SNP	15477
	TMT	892	AT	A			Indel	1951.73
	TMT	923	G	A	75	25	SNP	2519.77
	TMT	930	C	T	75	25	SNP	2432.77
	TMT	938	TA	T			Indel	2471.73
	TMT	942	A	T	79	21	SNP	24877
	TMT	946	GAACT	G			Indel	2558.73
	TMT	952	GACTA	G			Indel	2532.73
	TMT	1003	G	GTTCAATTTTGTTGTAGTGTATGATACACTACAACAAATGAAGTAGCATATCTACCATTACGCTACTTGAAAATTATGGCGTAATAATA			Indel	41173
	TMT	1049	G	A	74	26	SNP	2348.77
	TMT	1085	A	C			Indel	1704.77
	TMT	1090	CT	C			Indel	1667.73
	TMT	1098	A	AAAACCTATTTGTATTTGATCCTAAT			Indel	2324.73
	TMT	1124	G	C	67	33	SNP	2532.73
	TMT	1265	C	T	72	28	SNP	1268.77
	TMT	1293	C	T	71	29	SNP	16077
	TMT	1306	A	C	69	31	SNP	1725.77
	TMT	1319	G	A	68	32	SNP	1703.77

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
	TMT	1338	G	A	65	35	SNP	1085.73
	TMT	1347	A	T	65	35	SNP	2281.73
	TMT	1384	T	C	62	38	SNP	2641.77
	TMT	1440	G	GGTCTC			Indel	2812.73
	TMT	1466	C	T	63	37	SNP	2972.77
	TMT	1470	A	G	63	37	SNP	2864.77
	TMT	1490	T	C	64	36	SNP	2405.77
	TMT	1497	C	T	62	38	SNP	2543.77
	TMT	1504	A	T	62	38	SNP	2492.77
	TMT	1524	T	C	66	34	SNP	1937.77
	TMT	1531	T	C	70	30	SNP	1601.77
	TMT	1538	T	C	73	27	SNP	15277
	TMT	1633	A	C	70	30	SNP	1855.77
	TMT	1659	C	A	68	32	SNP	2118.77
	TMT	1668	T	C	69	31	SNP	2052.77
	TMT	1669	G	A	69	31	SNP	1835.77
	TMT	1703	C	A	73	27	SNP	2492.77
	TMT	1786	A	ATAG			Indel	1328.73
	TMT	2150	G	A	72	28	SNP	1616.77
	TMT	2187	C	T	69	31	SNP	1375.77
	TMT	2222	T	C	69	31	SNP	2028.77
	TMT	2244	AGTGATCG	A			Indel	2359.73
	TMT	2297	G	T	73	27	SNP	3358.77
	TMT	2360	A	C	70	30	SNP	7982.77
	TMT	2440	G	A	68	32	SNP	14217.77
	TMT	2457	T	A	70	30	SNP	11847.77

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
	TMT	2481	G	A	70	30	SNP	9333.77
	TMT	2503	C	T	68	32	SNP	8587.77
	TMT	2571	C	G	62	38	SNP	19877.77
	TMT	2580	T	G	63	37	SNP	18653.77
	TMT	2590	T	TC			Indel	18574.73
	TMT	2600	T	A	68	32	SNP	18151.77
	TMT	2601	A	C	69	31	SNP	18058.77
	TMT	2611	TGTG	T			Indel	18368.73
	TMT	2615	G	GCC			Indel	17997.73
	TMT	2828	G	T	53	47	SNP	11394.77
	TMT	2846	A	ACAAATCACAATCT ACCGTAAGAAAACA CTCTTAATGTGACA GTACTTTTTTTTTG ACAGTACATTAAAA GTATAATTTGCATC CGACAAGTTATTTA ATTTCAACTTTCCA ATCCTGAAAAATTT CACCGTTAATT			Indel	17112.73
	TMT	2876	A	G	58	42	SNP	9283.77
	TMT	3035	T	TC			Indel	3098.73
	TMT	3127	A	G	69	31	SNP	3901.77
	TMT	3132	T	C	70	30	SNP	39877
	TMT	3247	A	C	62	38	SNP	4459.77
	TMT	3272	T	C	65	35	SNP	38877

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
	TMT	3275	C	A	66	34	SNP	3494.77
	TMT	3321	AC	A			Indel	5843.73
	TMT	3329	C	T	66	34	SNP	5812.77
	TMT	3333	C	T	66	34	SNP	5742.77
	TMT	3334	C	A	65	35	SNP	5745.77
	TMT	3339	G	A	64	36	SNP	5753.77
	TMT	3375	C	T	66	34	SNP	4797.77
	TMT	3380	G	T	70	30	SNP	4129.77
	TMT	3575	T	A	62	38	SNP	5231.77
	TMT	3622	C	G	66	34	SNP	4574.77
	TMT	3625	CT	C			Indel	4529.73
	TMT	3636	C	T	68	32	SNP	3394.77
	TMT	3730	C	T	60	40	SNP	4737.77
	TMT	3832	A	T	64	36	SNP	1916.77
KP 472 2 SNP	*TMT*	3722	G	T	02	98	SNP	58357.77
	TMT	3832	A	T	0	100	SNP	27988.77
KP 697	*HPT*	156	T	C	55	55	SNP	5751.77
2 SNPs	*HPT*	157	G	C	55	55	SNP	5754.77
KP 704	*HPT*	156	T	C	50	50	SNP	6649.77
2 SNPs	*HPT*	157	G	C	50	50	SNP	6649.77
KP 729	*HPT*	87	G	T	01	99	SNP	14277.77

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
34 SNPs & 8 Indels	*HPT*	95	T	A	01	99	SNP	13167.73
	HPT	96	G	T	01	99	SNP	13334.77
	HPT	108	G	GAGTAGCT			Indel	88773
	HPT	174	T	TTC			Indel	1407.73
	HPT	177	G	T	07	93	SNP	1345.41
	HPT	208	G	C	0	100	SNP	9214.73
	HPT	215	G	A	41	69	SNP	10743.77
	HPT	222	G	T	01	99	SNP	14776.77
	HPT	250	A	C	69	31	SNP	6078.77
	HPT	302	T	C	03	97	SNP	56449.77
	HPT	346	A	T	01	99	SNP	54628.77
	HPT	351	T	G	01	99	SNP	54271.77
	HPT	387	T	TA			Indel	469573
	HPT	399	A	G	02	98	SNP	41647.77
	HPT	400	A	G	01	99	SNP	48469.77
	HPT	470	C	A	01	99	SNP	83477.77
	HPT	471	C	A	26	74	SNP	73282.77
	HPT	484	C	G	01	99	SNP	87089.77
	HPT	579	T	TGCTACTA			Indel	46189.73
	HPT	598	C	A	01	99	SNP	50505.77
	HPT	599	A	T	01	99	SNP	50037.77
	HPT	607	C	A	01	99	SNP	65206.77
	HPT	674	A	G	02	98	SNP	81152.77
	HPT	710	T	C	01	99	SNP	87243.77

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
	HPT	727	A	T	01	99	SNP	56799.73
	HPT	798	A	AGCAAGTAT			Indel	65747.73
	HPT	822	A	C	02	98	SNP	54498.77
	HPT	976	G	A	01	99	SNP	40033.77
	HPT	1038	TAA	T			Indel	38551.73
	HPT	1041	T	TGA			Indel	38026.73
	HPT	1098	C	T	01	99	SNP	51871.77
	HPT	1202	C	T	01	99	SNP	97838.77
	HPT	1217	C	T	02	98	SNP	86432.77
	HPT	1238	C	A	02	98	SNP	86365.77
	HPT	1271	T	C	01	99	SNP	51592.77
	HPT	1354	G	A	02	98	SNP	63524.77
	HPT	1372	A	C	01	99	SNP	63157.77
	HPT	1381	A	C	02	98	SNP	63888.77
	TMT	1106	ATG	A			Indel	7173
	TMT	3722	G	T	02	98	SNP	70485.77
	TMT	3832	A	T	0	100	SNP	31057.77
KP 734	*TC*	4258	CTTGAATCCT	C	04		Indel	15674.73
1 SNP & 1 Indel	*TC*	4597	G	C	0	100	SNP	238377
KP 759 1 Indel	*TMT*	1106	ATG	A	5		Indel	1235.73
KP 801	*HPT*	151	TA	T			Indel	1825.73

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
1 indel								
KP 806	*TMT*	217	T	C	01		SNP	21392.77
92 SNP & 26 Indels	*TMT*	229	T	C	0	100	SNP	24074.77
	TMT	279	T	C	0	100	SNP	19097.77
	TMT	342	T	A	01		SNP	14783.77
	TMT	424	T	C	0	100	SNP	174077
	TMT	466	AC	A			Indel	23868.73
	TMT	515	T	G	01		SNP	21111.77
	TMT	528	TGAAACTCAAAG CAAAA	T			Indel	17635.73
	TMT	559	C	T	0	100	SNP	22562.77
	TMT	598	G	C	0	100	SNP	29539.77
	TMT	630	C	G	0	100	SNP	28619.77
	TMT	705	T	TTAATACATAA			Indel	120573
	TMT	708	A	ATTATT			Indel	98373
	TMT	710	A	AGTCGGGAACCTC AATT			Indel	18765.73
	TMT	717	TA	T			Indel	16213.73
	TMT	722	C	T	03		SNP	18042.73
	TMT	779	A	G	0	100	SNP	28272.77
	TMT	795	A	G	0	100	SNP	27024.73
	TMT	806	C	T	0	100	SNP	26935.77
	TMT	856	C	T	20	80	SNP	40077

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
	TMT	858	C	T	10	90	SNP	4543.77
	TMT	860	C	T	04		SNP	7018.10
	TMT	862	C	T	0	100	SNP	125477
	TMT	864	C	T	0	100	SNP	13682.77
	TMT	875	G	A,GTA	06		SNP	13449.73
	TMT	887	A	G	0	100	SNP	24675.77
	TMT	892	AT	A			Indel	309773
	TMT	894	T	A	06		SNP	20229.77
	TMT	923	G	A	0	100	SNP	34995.77
	TMT	930	C	T	0	100	SNP	36626.77
	TMT	947	AACTAGACT	A			Indel	30625.73
	TMT	1002	T	TAATTACTCTA			Indel	1136.73
	TMT	1003	G	GTTCAATTTTCTTG TAGCGTGTGATACA ATACAACAAATGAA GTAGCATATCTATC GTTACGCTACTTGA AAACTATGGCGTAA TAATA			Indel	30224.73
	TMT	1049	G	A	0	100	SNP	31879.77
	TMT	1068	C	T	01		SNP	27989.77
	TMT	1085	A	C	0	100	SNP	23244.77
	TMT	1090	CT	C			Indel	220173
	TMT	1096	G	GTAAAACCTATTTG TATTTGATCC			Indel	8139.73
	TMT	1102	C	CTATA,CTATATA			Indel	5906.73

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
	TMT	1124	G	C	02		SNP	16026.73
	TMT	1186	T	C	0	100	SNP	18392.77
	TMT	1293	C	A	0	100	SNP	12049.77
	TMT	1319	G	A	0	100	SNP	11455.73
	TMT	1333	A	G	03		SNP	13312.77
	TMT	1337	T	TA			Indel	4315.73
	TMT	1338	G	A	01		SNP	10162.77
	TMT	1378	T	C	01		SNP	19031.77
	TMT	1384	T	C	0	100	SNP	18903.77
	TMT	1440	G	GTCTT			Indel	25193.73
	TMT	1466	C	T	0	100	SNP	21953.77
	TMT	1470	A	G	0	100	SNP	21005.77
	TMT	1490	T	C	0	100	SNP	18624.77
	TMT	1497	C	A	0	100	SNP	17862.77
	TMT	1518	T	C	01		SNP	17457.77
	TMT	1529	G	T	0	100	SNP	16116.77
	TMT	1534	T	A	0	100	SNP	14921.77
	TMT	1588	A	C	0	100	SNP	24362.77
	TMT	1629	A	AT			Indel	25741.73
	TMT	1633	A	C	01		SNP	24809.77
	TMT	1647	A	C	01		SNP	21616.77
	TMT	1659	C	A	0	100	SNP	18151.77
	TMT	1668	T	C	0	100	SNP	17035.77
	TMT	1669	G	A	01		SNP	13486.77
	TMT	1702	A	G	0	100	SNP	21505.73
	TMT	1746	GC	G			Indel	21352.73

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
	TMT	1786	A	ATAG			Indel	22282.73
	TMT	1790	C	T	01		SNP	25968.73
	TMT	1824	T	C	0	100	SNP	23886.77
	TMT	2125	C	A	0	100	SNP	18038.77
	TMT	2134	T	C	0	100	SNP	22663.77
	TMT	2145	T	C	0	100	SNP	21638.77
	TMT	2150	G	A	0	100	SNP	24972.77
	TMT	2155	TA	T			Indel	15837.73
	TMT	2185	G	A	0	100	SNP	18356.77
	TMT	2187	C	G	0	100	SNP	18377.77
	TMT	2203	A	ACCTTC			Indel	15085.73
	TMT	2212	C	T	0	100	SNP	26975.73
	TMT	2222	T	C	0	100	SNP	28167.77
	TMT	2225	T	C	0	100	SNP	26897.77
	TMT	2241	CATAGTG	C			Indel	22745.73
	TMT	2250	CG	C			Indel	25044.73
	TMT	2286	A	T	0	100	SNP	32179.77
	TMT	2360	A	C	0	100	SNP	26324.77
	TMT	2380	G	A	0	100	SNP	38325.77
	TMT	2393	T	C	0	100	SNP	36538.77
	TMT	2433	A	T	0	100	SNP	35511.77
	TMT	2435	G	C	0	100	SNP	41623.73
	TMT	2469	T	A	01		SNP	32425.77
	TMT	2481	G	A	0	100	SNP	30595.77
	TMT	2500	T	G	01		SNP	25559.73
	TMT	2503	C	T	01		SNP	240877

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
	TMT	2578	T	C	0	100	SNP	41832.77
	TMT	2601	A	C	01		SNP	45359.77
	TMT	2613	T	C	0	100	SNP	44494.77
	TMT	2763	G	T	0	100	SNP	47987.77
	TMT	2846	A	ACAAATCACAATCT ACTGTAAGAAAACA CTCTTAATGTGACA GTACTTTTTTTTGA CAGTACATTCAAAG TATAATTTGCATCC GACAAGTTATTTAA TTTCAACTTTCAAT CCTGAAAAATTTCA CCATTAATT			Indel	15135.73
	TMT	3009	G	T	0	100	SNP	38929.77
	TMT	3035	T	TC			Indel	30827.73
	TMT	3051	T	A	0	100	SNP	24227.77
	TMT	3089	C	G	0	100	SNP	20123.77
	TMT	3127	A	G	0	100	SNP	23614.77
	TMT	3132	T	C	0	100	SNP	26178.77
	TMT	3178	A	T	0	100	SNP	36527.77
	TMT	3189	A	T	0	100	SNP	37998.77
	TMT	3237	CAACTAACAAA	C			Indel	21738.73
	TMT	3265	C	T	0	100	SNP	21794.77
	TMT	3291	T	C	0	100	SNP	22799.77
	TMT	3329	C	T	0	100	SNP	27302.77

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
	TMT	3333	C	T	0	100	SNP	27472.77
	TMT	3334	C	A	01		SNP	24756.77
	TMT	3339	G	A	0	100	SNP	27424.77
	TMT	3375	C	T	0	100	SNP	24907.77
	TMT	3575	T	A	0	100	SNP	23916.77
	TMT	3622	C	G	0	100	SNP	23686.77
	TMT	3625	CT	C			Indel	23612.73
	TMT	3636	C	T	0	100	SNP	175677
	TMT	3701	A	C	0	100	SNP	22402.77
	TMT	3832	A	T	0	100	SNP	8747.77
KP 1170 74 SNP & 18 Indels	*TMT*	142	C	T	0	99	SNP	8244.77
	TMT	151	TAC	T			Indel	11506.73
	TMT	155	C	CCG			Indel	11826.73
	TMT	222	C	T	0	99	SNP	12861.77
	TMT	229	T	C	0	98	SNP	12646.77
	TMT	264	C	T	0	100	SNP	10604.77
	TMT	466	AC	A			Indel	10838.73
	TMT	504	A	G	0	100	SNP	15023.77
	TMT	527	T	C	0	100	SNP	166577
	TMT	533	C	T	0	99	SNP	18475.77
	TMT	538	A	T	0	100	SNP	18259.77
	TMT	563	T	TA			Indel	13219.73
	TMT	664	G	C	0	99	SNP	15631.77

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
	TMT	701	A	G	0	99	SNP	205577
	TMT	854	C	T	15	85	SNP	8528.77
	TMT	864	CGT	C			Indel	8467.73
	TMT	930	C	T	0	100	SNP	22098.77
	TMT	1049	G	A	0	100	SNP	28825.77
	TMT	1106	A	ATG			Indel	5415.73
	TMT	1338	GA	G			Indel	9199.73
	TMT	1384	T	C	0	99	SNP	13803.77
	TMT	1435	C	T	0	100	SNP	15935.77
	TMT	1440	G	GGTCTC			Indel	17237.73
	TMT	1451	G	A	0	100	SNP	16193.77
	TMT	1453	C	T	0	100	SNP	16202.77
	TMT	1470	A	G	0	100	SNP	16366.77
	TMT	1490	T	C	0	100	SNP	118477
	TMT	1497	C	T	0	99	SNP	11237.77
	TMT	1524	T	C	0	100	SNP	9528.77
	TMT	1529	G	T	0	100	SNP	99177
	TMT	1534	T	A	0	99	SNP	9262.77
	TMT	1538	T	C	01	98	SNP	8884.77
	TMT	1615	C	CTA			Indel	8687.73
	TMT	1633	A	C	04	96	SNP	114577
	TMT	1640	C	T	0	100	SNP	11135.77
	TMT	1660	T	C	0	100	SNP	10305.77
	TMT	1668	T	C	0	100	SNP	10777.77
	TMT	1672	C	T	0	99	SNP	104077
	TMT	1709	A	G	0	99	SNP	107577

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
	TMT	1711	G	C	0	100	SNP	108477
	TMT	1734	T	C	0	100	SNP	9759.77
	TMT	1736	A	G	0	100	SNP	9478.77
	TMT	1762	C	T	0	99	SNP	12519.77
	TMT	1768	G	T	0	100	SNP	126477
	TMT	1790	C	T	0	99	SNP	11808.77
	TMT	1871	T	A	0	99	SNP	6467.77
	TMT	2103	C	A	0	100	SNP	90577
	TMT	2150	G	A	01	99	SNP	14075.77
	TMT	2155	TA	T			Indel	9566.73
	TMT	2187	C	T	0	100	SNP	14414.77
	TMT	2211	C	T	0	99	SNP	23444.77
	TMT	2222	T	C	01	99	SNP	25509.77
	TMT	2244	AGTGATCG	A			Indel	27509.73
	TMT	2288	TTAAA	T			Indel	46799.73
	TMT	2360	A	C	0	99	SNP	375677
	TMT	2486	C	T	0	100	SNP	34869.77
	TMT	2503	C	T	01	99	SNP	34586.77
	TMT	2571	C	G	0	99	SNP	53021.73
	TMT	2601	A	C	0	100	SNP	54101.77
	TMT	2604	GACAA	G			Indel	47326.73
	TMT	2613	T	C	01	99	SNP	47244.77
	TMT	2782	G	A	0	100	SNP	372377
	TMT	2808	G	C	0	99	SNP	363077
	TMT	2829	T	C	0	99	SNP	31983.73
	TMT	2846	A	ACAAATCACAATCT			Indel	13678.73

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
				ACTGTAAGAAAACA CTCTTAATGTGACA GTACTTTTTTTTGA CAGTATATTAAAAG TATAATTTGCATCC GACAAGTTATTTAA TTTCAACTTTCAAT CCTGAAAAATTTCA CCATTAATT				
	TMT	2928	G	A	01	99	SNP	11368.77
	TMT	3035	T	TC			Indel	12965.73
	TMT	3041	G	A	01	96	SNP	22792.73
	TMT	3085	C	T	01	99	SNP	13575.77
	TMT	3127	A	G	0	100	SNP	20465.77
	TMT	3132	T	C	0	100	SNP	22523.77
	TMT	3153	A	G	01	99	SNP	27691.77
	TMT	3156	TA	T			Indel	24911.73
	TMT	3178	A	C	01	99	SNP	31769.77
	TMT	3189	A	T	0	99	SNP	31616.77
	TMT	3202	T	A	0	100	SNP	31602.73
	TMT	3240	CTAACAAA	C			Indel	20297.73
	TMT	3275	C	A	0	100	SNP	13652.77
	TMT	3329	C	T	01	99	SNP	19403.77
	TMT	3333	C	T	0	100	SNP	23232.77
	TMT	3334	C	A	0	100	SNP	23056.77
	TMT	3339	G	A	0	100	SNP	23371.77

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
	TMT	3375	C	T	0	100	SNP	24155.77
	TMT	3380	G	A	0	100	SNP	23612.77
	TMT	3438	T	A	0	100	SNP	18321.77
	TMT	3569	C	G	0	99	SNP	21301.77
	TMT	3622	C	G	0	99	SNP	16522.77
	TMT	3625	CT	C			Indel	14999.73
	TMT	3636	C	T	0	100	SNP	14055.77
	TMT	3701	A	C	0	100	SNP	16344.77
	TMT	3832	A	T	0	98	SNP	7589.77
	TMT	3898	T	C	0	100	SNP	4444.77
KP 1185 12 SNP & 1 Indel	*TMT*	342	T	A	03	93	SNP	1061.77
	TMT	1102	C	CTATATA			Indel	3075
	TMT	2360	A	C	0	100	SNP	1378.77
	TMT	2380	G	A	0	100	SNP	1933.77
	TMT	2393	T	C	0	100	SNP	1742.77
	TMT	2433	A	T	0	100	SNP	1775.77
	TMT	2435	G	C	0	100	SNP	1775.77
	TMT	2500	T	G	0	100	SNP	9077
	TMT	2578	T	C	0	98	SNP	1911.77
	TMT	2601	A	C	0	98	SNP	1979.77
	TMT	2613	T	C	0	98	SNP	1903.77
	TMT	2763	G	T	0	100	SNP	1491.77
	TMT	3575	T	A	03	97	SNP	1283.77

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
KP 1190 48 SNP & 12 Indels	*TMT*	217	T	C	0	98	SNP	1789.77
	TMT	229	T	C	0	100	SNP	3219.84
	TMT	279	T	C	02	97	SNP	2526.77
	TMT	342	T	A	01	97	SNP	2491.77
	TMT	424	T	C	0	100	SNP	8677
	TMT	466	AC	A			Indel	1173.73
	TMT	515	T	G	0	100	SNP	1234.77
	TMT	528	TGAAACTCAAAG CAAAA	T			Indel	1155.73
	TMT	630	C	G	03	97	SNP	1511.77
	TMT	779	A	G	0	100	SNP	1491.77
	TMT	795	A	G	0	97	SNP	1449.77
	TMT	806	C	T	0	95	SNP	1475.77
	TMT	852	TGCGC	T			Indel	1277
	TMT	875	G	GTA			Indel	1022.73
	TMT	887	A	G	0	100	SNP	1321.77
	TMT	892	AT	A			Indel	1449.73
	TMT	894	T	A	13	88	SNP	1413.77
	TMT	923	G	A	0	100	SNP	1944.77
	TMT	930	C	T	0	100	SNP	2025.77
	TMT	947	AACTAGACT	A			Indel	1979.73
	TMT	1002	T	TAATTACTCTA			Indel	1637.73
	TMT	1003	G	GTTCAATTTTCTTG TAGCGTGTGATACA			Indel	1637.73

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
				ATACAACAAATGAA GTAGCATATCTATC GTTACGCTACTTGA AAACTATGGCGTAA TAATA				
	TMT	1440	G	GTCTT			Indel	1201.73
	TMT	1633	A	C	10	90	SNP	1032.77
	TMT	1647	A	C	0	95	SNP	1006.77
	TMT	1659	C	A	0	100	SNP	1006.77
	TMT	1702	A	G	0	97	SNP	1178.77
	TMT	1824	T	C	0	97	SNP	11477
	TMT	2241	CATAGTG	C			Indel	1042.73
	TMT	2250	CG	C			Indel	1078.73
	TMT	2286	A	T	0	100	SNP	1268.77
	TMT	2360	A	C	0	100	SNP	2176.77
	TMT	2380	G	A	0	100	SNP	3038.77
	TMT	2393	T	C	0	100	SNP	2772.77
	TMT	2433	A	T	0	100	SNP	26777
	TMT	2435	G	C	0	95	SNP	26777
	TMT	2469	T	A	0	100	SNP	2324.77
	TMT	2481	G	A	0	100	SNP	2174.77
	TMT	2500	T	G	0	100	SNP	1861.77
	TMT	2503	C	T	0	97	SNP	1726.77
	TMT	2578	T	C	0	99	SNP	3661.77
	TMT	2601	A	C	01	99	SNP	4053.77
	TMT	2613	T	C	01	99	SNP	3895.77

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
	TMT	2763	G	T	0	100	SNP	3707.77
	TMT	3009	G	T	0	100	SNP	1152.77
	TMT	3127	A	G	0	100	SNP	1262.77
	TMT	3132	T	C	0	100	SNP	1289.77
	TMT	3178	A	T	0	95	SNP	1658.77
	TMT	3189	A	T	0	95	SNP	1567.77
	TMT	3329	C	T	0	97	SNP	1767.77
	TMT	3333	C	T	0	100	SNP	1726.77
	TMT	3334	C	A	14	96	SNP	1726.77
	TMT	3339	G	A	0	100	SNP	1719.77
	TMT	3375	C	T	0	98	SNP	1679.77
	TMT	3575	T	A	0	98	SNP	1789.77
	TMT	3622	C	G	0	91	SNP	1436.77
	TMT	3625	CT	C			Indel	1399.73
	TMT	3636	C	T	0	100	SNP	1307.77
	TMT	3701	A	C	0	100	SNP	1567.77
	TMT	3832	A	T	0	96	SNP	9077
KP 1191 10 SNP & 1 Indel	*TMT*	515	T	G	0	94	SNP	6777
	TMT	2250	CG	C			Indel	1007.73
	TMT	2380	G	A	0	96	SNP	1187.77
	TMT	2393	T	C	0	100	SNP	1112.77
	TMT	2433	A	T	0	100	SNP	1096.77
	TMT	2435	G	C	0	100	SNP	1096.77
	TMT	2578	T	C	0	100	SNP	1361.77

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
	TMT	2601	A	C	0	100	SNP	1537.77
	TMT	2613	T	C	0	100	SNP	1493.77
	TMT	2763	G	T	0	98	SNP	1581.77
	TMT	3009	G	T	0	100	SNP	7177
KP 1219 31 SNP & 5 Indels	*HPT*	243	T	C	22	78	SNP	1738.74
	HPT	262	C	A	25	74	SNP	1952.77
	HPT	330	T	G	13	85	SNP	29377
	HPT	346	A	T	07	92	SNP	3768.77
	HPT	351	T	G	09	91	SNP	3456.77
	HPT	390	G	A	13	87	SNP	2829.77
	HPT	399	A	G	11	89	SNP	2998.77
	HPT	400	A	G	09	90	SNP	2921.77
	HPT	445	C	A	08	91	SNP	2531.77
	HPT	447	A	G	07	91	SNP	2531.77
	HPT	470	C	A	08	92	SNP	30677
	HPT	471	C	A	10	89	SNP	31379
	HPT	484	C	G	07	89	SNP	3139.77
	HPT	537	C	G	08	89	SNP	4693.77
	HPT	541	GT	G			Indel	4646.73
	HPT	598	C	A	13	85	SNP	5101.96
	HPT	599	A	C	13	87	SNP	5261.56
	HPT	603	T	A	13	86	SNP	5556.77
	HPT	607	C	A	11	89	SNP	5854.78

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
	HPT	710	T	C	10	88	SNP	7714.77
	HPT	712	T	G	12	87	SNP	8049.77
	HPT	727	AAAAGGAG	A			Indel	4584.73
	HPT	735	TCAAACAG	T			Indel	4631.73
	HPT	798	A	AGCAAGTAT			Indel	8259.73
	HPT	937	A	T	5	93	SNP	9132.77
	HPT	976	G	A	8	92	SNP	8748.77
	HPT	1029	CG ATATGATTAATTT GACCACAATAGT	C			Indel	5288.73
	HPT	1031	CAAACG	A			Indel	5520.73
	HPT	1083	C	T	8	92	SNP	10526.77
	HPT	1098	C	T	8	90	SNP	9811.77
	HPT	1146	T	C	9	89	SNP	8717.77
	HPT	1202	C	T	11	89	SNP	7330.77
	HPT	1244	C	A	11	89	SNP	7223.77
	HPT	1271	T	C	9	88	SNP	5167.77
	HPT	1312	G	C	5	94	SNP	5463.77
	TMT	3722	G	T	1	99	SNP	58477.77
	TMT	3832	A	T	0	98	SNP	27549.77
KP 1227 55 SNP & 19 Indels	*TMT*	154	G	C	1	98	SNP	25401.77
	TMT	155	C	G	0	99	SNP	25479.77

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
	TMT	178	G	C	0	97	SNP	30790.77
	TMT	195	C	A	1	98	SNP	27262.77
	TMT	217	T	C	1	96	SNP	27195.77
	TMT	229	T	C	0	99	SNP	30381.77
	TMT	468	C	T	0	97	SNP	23983.77
	TMT	533	C	T	1	99	SNP	39766.77
	TMT	538	A	T	0	99	SNP	39428.77
	TMT	705	T	TTAATACATAA			Indel	18267.73
	TMT	708	A	ATTATT			Indel	16541.73
	TMT	710	A	AGTCGGGAACCTC AATT			Indel	30603.73
	TMT	717	TA	T			Indel	21487.73
	TMT	779	A	G	0	99	SNP	39355.77
	TMT	858	CGCGCGCGT	C			Indel	17784.73
	TMT	875	G	A	9	91	SNP	27288.73
	TMT	877	A	AT			Indel	17634.73
	TMT	892	ATTT	A			Indel	41593.73
	TMT	895	T	TAA			Indel	35463.73
	TMT	908	A	T	0	99	SNP	41156.77
	TMT	917	C	T	1	99	SNP	39289.77
	TMT	923	G	A	0	98	SNP	44979.77
	TMT	930	C	T	0	99	SNP	50801.77
	TMT	947	AACTAGACT	A			Indel	47281.73
	TMT	1002	T	TAATTACTCTA GTTCAATTTTGTTG			Indel	2677.73
	TMT	1003	G	TAGCGTATGATACA			Indel	21370.73

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
				CTACAACAAATGAA GTAGCATATCTATC ATTACGCTACTTGA AAACTATGGCGTAA TAATA				
	TMT	1018	A	G	43	56	SNP	28160.73
	TMT	1049	G	A	0	99	SNP	31820.77
	TMT	1060	G	A	1	99	SNP	52741.77
	TMT	1085	A	C	0	99	SNP	40957.77
	TMT	1090	CT	C			Indel	37748.73
	TMT	1096	G	GTAAAACCTATTTG TATTTGATCC			Indel	10375.73
	TMT	1102	C	CTATATA			Indel	7200.73
	TMT	1117	T	TGC			Indel	6121.73
	TMT	1124	G	C,GTC			Indel	25899.73
	TMT	1186	T	C	0	98	SNP	25735.77
	TMT	1287	C	A	1	98	SNP	18469.77
	TMT	1293	C	A	0	99	SNP	17945.77
	TMT	1319	G	A	0	100	SNP	12301.77
	TMT	1337	T	TA			Indel	4167.73
	TMT	1338	G	A	0	99	SNP	10662.73
	TMT	1378	T	C	1	98	SNP	15224.77
	TMT	1384	T	C	0	98	SNP	16743.77
	TMT	1403	A	G	0	99	SNP	18864.77
	TMT	1405	A	G	0	99	SNP	18871.77
	TMT	1408	T	A	0	99	SNP	19681.77

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
	TMT	1417	T	C	0	99	SNP	23416.77
	TMT	1435	C	T	1	99	SNP	21913.77
	TMT	1440	G	GTCTT			Indel	27895.73
	TMT	1470	A	G	0	99	SNP	30002.77
	TMT	1490	T	C	0	99	SNP	23213.77
	TMT	1497	C	A	1	99	SNP	22520.77
	TMT	1518	T	C	0	99	SNP	20618.73
	TMT	1529	G	T	1	97	SNP	12843.77
	TMT	1532	G	A	1	98	SNP	12640.77
	TMT	1534	T	A	0	100	SNP	14914.73
	TMT	1629	A	AT			Indel	19772.73
	TMT	1633	A	C	5	94	SNP	20461.77
	TMT	1635	G	T	1	96	SNP	22652.77
	TMT	1659	C	A	0	100	SNP	19771.77
	TMT	1668	T	C	0	99	SNP	19320.77
	TMT	1669	G	A	0	98	SNP	17488.77
	TMT	1709	A	G	0	99	SNP	19728.77
	TMT	1711	G	C	0	99	SNP	19268.77
	TMT	1734	T	C	0	99	SNP	16973.77
	TMT	1736	A	G	0	99	SNP	17063.77
	TMT	1762	C	T	0	98	SNP	20851.77
	TMT	1768	G	T	0	99	SNP	20901.77
	TMT	1790	C	T	0	100	SNP	19263.77
	TMT	1825	G	A	0	98	SNP	15957.77
	TMT	1898	G	A	0	98	SNP	13085.77
	TMT	2018	G	T	0	100	SNP	12583.77

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
	TMT	2084	C	A	0	100	SNP	10498.77
	TMT	2377	C	A	0	99	SNP	53413.77
KP 159657 SNP & 26 Indels	*TMT*	466	AC	A			Indel	59981.73
	TMT	528	TGAAACTCAAAG CAAAA	T			Indel	23711.73
	TMT	705	T	TTAATACATAA			Indel	16776.73
	TMT	708	A	ATTATT AGTCGGGAACCTC			Indel	17148.73
	TMT	710	A	AATT			Indel	36354.73
	TMT	717	TA	T			Indel	19932.73
	TMT	852	TGCGC	T			Indel	1450.73
	TMT	856	C	T	17	82	SNP	9007.77
	TMT	858	C	T	7	92	SNP	9533.77
	TMT	860	C	T	2	98	SNP	15741.77
	TMT	862	C	T	2	98	SNP	18725.77
	TMT	864	C	T	2	97	SNP	22561.73
	TMT	887	A	G	0	100	SNP	46684.77
	TMT	892	AT	A			Indel	42166.73
	TMT	923	G	A	0	99	SNP	42437.77
	TMT	930	C	T	0	99	SNP	43640.77
	TMT	947	AACTAGACT	A			Indel	28810.73
	TMT	1002	T	TAATTACTCTA GTTCAATTTTCTTG			Indel	2073.73
	TMT	1003	G	TAGCGTGTGATACA			Indel	13074.73

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
				ATACAACAAATGAA GTAGCATATCTATC GTTACGCTACTTGA AAACTATGGCGTAA TAATA				
	TMT	1049	G	A	0	99	SNP	40728.77
	TMT	1068	C	T	0	99	SNP	42365.77
	TMT	1085	A	C	0	99	SNP	38127.77
	TMT	1090	CT	C			Indel	32173.73
	TMT	1096	G	GTAAAACCTATTTG TATTTGATCC			Indel	37927.73
	TMT	1102	C	CTATA,CTATATA			Indel	10652.73
	TMT	1124	G	C	3	96	SNP	35687.73
	TMT	1186	T	C	0	98	SNP	89822.77
	TMT	1293	C	A	1	98	SNP	49182.77
	TMT	1319	G	A	0	99	SNP	44850.77
	TMT	1333	A	G	2	96	SNP	53780.77
	TMT	1337	T	TA			Indel	17156.73
	TMT	1338	G	A	1	98	SNP	38355.77
	TMT	1378	T	C	0	100	SNP	67944.77
	TMT	1384	T	C	0	99	SNP	64714.77
	TMT	1440	G	GTCTT			Indel	77917.73
	TMT	1466	C	T	0	99	SNP	77273.77
	TMT	1470	A	G	0	99	SNP	75604.77
	TMT	1490	T	C	0	99	SNP	77180.77
	TMT	1497	C	A	0	99	SNP	78427.77

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
	TMT	1518	T	C	0	99	SNP	98035.77
	TMT	1529	G	T	0	99	SNP	99274.77
	TMT	1534	T	A	0	99	SNP	93258.77
	TMT	1588	A	C	0	99	SNP	76373.77
	TMT	1629	A	AT			Indel	59997.73
	TMT	1633	A	C	2	97	SNP	59308.77
	TMT	1647	A	C	0	99	SNP	48841.77
	TMT	1659	C	A	1	99	SNP	29478.77
	TMT	1668	T	C	0	99	SNP	36358.77
	TMT	1669	G	A	0	99	SNP	34073.77
	TMT	1702	A	G	0	98	SNP	48996.77
	TMT	1746	GC	G			Indel	37070.73
	TMT	1786	A	ATAG			Indel	42747.73
	TMT	1790	C	T	0	99	SNP	42081.77
	TMT	1824	T	C	0	99	SNP	39577.77
	TMT	2125	C	A	0	99	SNP	54541.77
	TMT	2134	T	C	0	99	SNP	71069.77
	TMT	2145	T	C	0	99	SNP	73530.77
	TMT	2150	G	A	0	99	SNP	74354.77
	TMT	2155	TA	T			Indel	47513.73
	TMT	2203	A	ACCTTC			Indel	37175.73
	TMT	2212	C	T	0	99	SNP	71577.73
	TMT	2241	CATAGTG	C			Indel	41975.73
	TMT	2250	CG	C			Indel	45021.73
	TMT	2481	G	A	0	99	SNP	39818.77
	TMT	2500	T	G	0	99	SNP	35756.77

Sample	Gene	Position	Reference allele	Alternative allele	Reference allele frequency	Alternative allele frequency	Type	Quality score
	TMT	2503	C	T	0	99	SNP	33676.77
	TMT	3035	T	TC			Indel	88594.73
	TMT	3089	C	G	0	99	SNP	49402.77
	TMT	3127	A	G	0	99	SNP	69570.77
	TMT	3132	T	C	0	99	SNP	74497.77
	TMT	3237	CAACTAACAAA	C			Indel	24807.73
	TMT	3265	C	T	0	99	SNP	29782.77
	TMT	3329	C	T	0	99	SNP	49637.77
	TMT	3333	C	T	0	99	SNP	43039.77
	TMT	3334	C	A	0	99	SNP	43347.77
	TMT	3339	G	A	0	99	SNP	50411.77
	TMT	3375	C	T	0	99	SNP	40974.77
	TMT	3575	T	A	0	99	SNP	55991.77
	TMT	3622	C	G	0	99	SNP	45677.77
	TMT	3625	CT	C			Indel	45383.73
	TMT	3636	C	T	0	99	SNP	41718.77
	TMT	3701	A	C	0	99	SNP	44984.77
	TMT	3832	A	T	0	99	SNP	76669.77

Reference (Appendix 4)

Garcia I, Rodgers M, Pepin R, Hsieh TF and Matringe M 1999. Characterization and subcellular compartmentation of recombinant 4-hydroxyphenylpyruvate dioxygenase from *Arabidopsis* in transgenic tobacco. *Journal of Plant Physiology* 1507-1516

García-moreno MJ, Fernández-martínez JM, Velasco L and Pérez-vich B 2012. Genetic basis of unstable expression of high gamma-tocopherol content in sunflower seeds. *BMC Plant Biology* 1-14

Hass CG, Tang S, Leonard S, Traber MG, Miller JF and Knapp SJ 2006. Three non-allelic epistatically interacting methyltransferase mutations produce novel tocopherol (vitamin E) profiles in sunflower. *Theoretical and Applied Genetics* 767-782

Hunter SC and Cahoon EB 2007. Enhancing Vitamin E in Oilseeds: Unraveling tocopherol and tocotrienol biosynthesis. *Lipids* 97-108

Kanwischer M, Porfirova S, Bergmuller E and Dormann P 2005. Alterations in tocopherol cyclase activity in transgenic and mutant plants of *Arabidopsis* affect tocopherol content, tocopherol composition, and oxidative stress. *Journal of Plant Physiology* 713-723

Maeda H, Sage TL, Welti R and DellaPenna D 2008. Tocopherols modulate extraplastidic polyunsaturated fatty acid metabolism in *Arabidopsis* at low temperature. *The Plant Cell* 452-470

Motohashi R, Ito T, Kobayashi M, Taji T, Nagata N, Asami T, Yoshida S, Yamaguchi-Shinozaki K and Shinozaki K 2003. Functional analysis of the 37kDa inner envelope membrane polypeptide in chloroplast biogenesis using a Ds-tagged *Arabidopsis* pale-green mutant. *The Plant Journal* 719-731

Ren W, Zhao L, Zhang L, Wang Y, Cui L, Tang Y, Sun X and Tang K 2011. Molecular analysis of a homogentisate phytyltransferase gene from *Lactuca sativa L. Molecular Biology Reports* 1813-1819

Chapter 5

Conclusion and recommendations

In this study 104 sunflower accessions were phenotyped for their linoleic and oleic acids levels, as well as for their tocopheol content. The results obtained indicate that about 90% of the 104 sunflower accessions tested had high levels of linoleic acid. The sunflower accession KP 417 consisted of high levels of both oleic and linoleic acid, higher than the calibrated value. According to the manufacturers of the Perten NIR and Spinlock NMR systems, the moisture content of seeds should be 10% or below, but KP 417 and KP 1226 consisted of 14% moisture content. Moisture content > 10% is considered irrelevant as it will influence the fatty acid content and give inaccurate results, thus indicating the importance of moisture content in seeds. Thus the results for KP 417 and KP 1226 were not considered as these accessions should be dried to a moisture content < 10% and reanalysed. Compared to oleic acid, higher levels of linoleic acid were found for the 104 sunflower accessions, proving that sunflower has higher levels of polyunsaturated fatty acid than monounsaturated fatty acid. The 20 commercial lines tested had low levels of oleic acid (< 25%) but higher levels of total oil, which argumentably states/supports that these commercial lines were bred for total oil yield.

All four tocopherol derivatives were successfully extracted from 104 sunflower accessions and tested on the GC MS/MS. The results obtained indicated that *α* tocopherol was the predominant tocopherol and had a correlation with total tocopherols (r = 0.89), thus indicating the importance of this specific tocopherol derivative. A positive correlation was also observed between *α* and total tocopherol, *β* and total tocopherol, *γ* and total tocopherol and *δ* and total tocopherol. Analysis also indicated a positive correlation between oleic acid and *γ* tocopherol, and oleic acid and *δ* tocopherol but no significance was observed among oleic acid, linoleic acid and all four tocopherol derivatives. Therefore, according to the results obtained in this study, tocopherol concentrations do not seem to be related to linoleic or oleic acid, thus breeding can be done for both tocopherol content and fatty acid concentrations at the same time.

Further genetic investigation for the vitamin E genes responsible for the production of all four tocopherol derivatives was done using next generation sequencing technologies. The vitamin E biosynthetic gene homologues were successfully identified isolated and characterized using the sunflower draft genome (2012) and public bioinformatic available resources. The information obtained from the bioinformatic available platforms were used to find the fatty acid reference pathway. This reference pathway was further analysed to construct the biosynthetic pathway of vitamin E. Custom designed primers were developed for amplifying the four targeted genes ρ-hydroxyphenylpyruvate dioxygenase (*HPPD*), homogentisate phytyltransferase (*HPT*), *γ*-tocopherol methyltransferase (*TMT***)**, and tocopherol/tocotrienol cyclase (*TC*). No sunflower gene sequences were available for *HPT* and *HPPD* but from the data obtained in this project a sunflower gene database was successfully created. The BLASTx results obtained indicate that the gene homogentisate phytylprenyltransferase (*HPPT*) was found instead of the targeted gene *HPPD* in the sunflower accessions analysed. The gene *HPPT* is a family member of *HPPD*, which is involved in tocopherol/tocotrienol biosynthesis depending on the type of antioxidant produced. Sequence analysis revealed that this gene consisted of a large gap (397 bp), which resulted in false positive SNPs over regions with no sequence coverage. Therefore this gene was excluded from this project.

From the 104 sunflower accessions analysed in chapter 3 a total of 23 accessions were selected and categorized in six different groups for sequencing (chapter 4). The six groups were formed to make mutation and expression predictions based on the phenotypic characteristics (chapter 3) and the genes involved in the production of the four tocopherol derivatives on the biosynthetic pathway of vitamin E. The sequencing results obtained indicated that polymorphisms were found for > 70% of the 23 accessions, with SNPs discovered in five of the six phenotypically described groups. A total number of 489 SNPs and 145 indels were found for the accessions tested. *TC* was conserved for > 98% of the accessions and consisted of a single allele but was found in the intron region of the gene. This might be due to the low production of *δ* tocopherol that was found across all 23 accessions. The

highest number of SNPs was found for group one (165 SNP's for *TMT* and 32 for *HPT*) and the least for group four (2 SNP's for *HPT*). *TMT* had the highest number of SNPs and consisted of 424 SNPs in total whereas *HPT* only had 64 SNPs in total. However, only 112 *TMT* and 27 *HPT* SNPs were located in the coding region of the genes. Of these, only 139 SNPs could cause translational changes and may play a role in protein structure and thus function. These mutations can be used to manipulate the genes of interest on the biosynthetic pathway of vitamin E to breed for traits of interest in future.

The gene 2-Methyl-6-prenylbenzoquinol methyltransferase (*MSBQ*) is situated in the middle of the biosynthetic pathway of vitamin E and could not be amplified during this study. This gene seems to play a crucial role in the synthesis of *α*, *β*, *γ* and *δ* tocopherol. Therefore it is recommended that amplification and sequencing should follow for *MSBQ*. This can unlock information needed to better understand the synthesis of the four tocopherol derivatives. Tang et al (2006) have identified and isolated two *MSBQ-MT* paralogs from sunflower (*MT-1* and *MT-2*). These alleles were resequenced from wildtype (*m*+ *m*+) and matant inbred lines (*m m*) and uncovered a non-lethal knockout mutant (*m*) of *MT-1* and a cryptic codominant mutant (*d*) in the wild type x mutant F_2 population (Tang et al 2006). The *d* mutant significantly increased *β* tocopherol in the *m m* individuals but had no effect on the *m*+ *m*+ individuals. The MT-2 mutants (*d*+ *d*+) were better transcribed in the wildtype seeds and leaves compared to the mutant homozygotes (*d d*). A double mutant was produced (*m m d d*) and enhanced *α* tocopherol (24-25%) and *β* tocopherol (55-74%) while the wildtype produced 96% *α* and 4% *β* tocopherol (Thang et al 2006). This indicates that the *MT-2* mutant has a major effect on the synthesis of tocopherol but does not have any effect on the plastoquinone production, as plastoquinone synthesis is very important for normal plant growth and development. The work of Tang et al (2006) therefore indicates that genes of interest can be targeted and manipulated to improve vitamin E content in sunflower using known mutations. It is also recommended that re-sequencing and amplification be done for *HPPD* because this enzyme is found central to the biosynthesis of all quinoid compounds such as plastoquinones and tocopherols that are both essential molecules in plants.